(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
27 December 2001 (27.12.2001)

PCT

(10) International Publication Number
WO 01/98508 A2

(51) International Patent Classification[7]: C12N 15/64

(21) International Application Number: PCT/US01/20015

(22) International Filing Date: 21 June 2001 (21.06.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/213,190	21 June 2000 (21.06.2000)	US
60/213,670	21 June 2000 (21.06.2000)	US
60/213,671	21 June 2000 (21.06.2000)	US
60/214,028	21 June 2000 (21.06.2000)	US
60/213,201	21 June 2000 (21.06.2000)	US
60/223,123	7 August 2000 (07.08.2000)	US
60/266,046	1 February 2001 (01.02.2001)	US
60/280,705	29 March 2001 (29.03.2001)	US
60/282,668	9 April 2001 (09.04.2001)	US
60/282,701	9 April 2001 (09.04.2001)	US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:
US	60/213,201 (CON)
Filed on	21 June 2000 (21.06.2000)
US	60/213,190 (CON)
Filed on	21 June 2000 (21.06.2000)
US	60/213,670 (CON)
Filed on	21 June 2000 (21.06.2000)
US	60/213,671 (CON)
Filed on	21 June 2000 (21.06.2000)
US	60/214,028 (CON)
Filed on	21 June 2000 (21.06.2000)
US	60/223,123 (CON)
Filed on	7 August 2000 (07.08.2000)
US	60/266,046 (CON)
Filed on	1 February 2001 (01.02.2001)
US	60/280,705 (CON)
Filed on	29 March 2001 (29.03.2001)
US	60/282,668 (CON)
Filed on	9 April 2001 (09.04.2001)

US 60/282,701 (CON)
Filed on 9 April 2001 (09.04.2001)

(71) Applicant (for all designated States except US): DELTA-GEN, INC. [US/US]; 740 Bay Road, Redwood City, CA 94063 (US).

(72) Inventors; and
(75) Inventors/Applicants (for US only): BRENNAN, Thomas, J. [US/US]; 325 Rockwood Drive, South San Francisco, CA 94080 (US). LEVITEN, Michael, W. [US/US]; 3166 Bryant Street, Palo Alto, CA 94306 (US). KLEIN, Robert [US/US]; 1044 Webster Street, Palo Alto, CA 94301 (US). MATTHEWS, William [US/US]; 560 Summit Springs Road, Woodside, CA 94062 (US). MOORE, Mark [US/US]; 880 Blandford Avenue, Redwood City, CA 94062 (US).

(74) Agents: LAPIZ, Mariette, A. et al.; Deltagen, Inc., 1003 Hamilton Avenue, Menlo PArk, CA 94025 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: TRANSGENIC MICE CONTAINING TARGETED GENE DISRUPTIONS

(57) Abstract: The present invention relates to transgenic animals, as well as compositions and methods relating to the characterization of gene function. Specifically, the present invention provides transgenic mice comprising mutations in a *TARGET* gene. Such transgenic mice are useful as models for disease and for identifying agents that modulate gene expression and gene function, and as potential treatments for various disease states and disease conditions.

-1-

TRANSGENIC MICE CONTAINING TARGETED GENE DISRUPTIONS

Field of the Invention

The present invention relates to transgenic animals, compositions and methods relating to the characterization of gene function.

Background of the Invention

Experimental animal models are important tools for understanding the role of genes. More particularly, the ability to develop animals with specific genes altered or inactivated has been invaluable to the study of gene function, and has lead to unexpected discoveries of a gene and/or mechanisms responsible for disease with similar manifestations in humans. These genetically engineered animals are also useful for identifying and testing therapeutic treatments for a variety of diseases and disorders.

The identification of the function of numerous genes has been useful in ascertaining the roles of these genes in disease. Because of the high level of homology between humans and mice, for example, it is possible to define the function of individual human genes by making targeted germline mutations in selected genes in the animal. The phenotype of the resulting mutant animal can be used to help define the phenotype in humans.

Several interesting genes have recently been discovered belonging to various families encoding G-protein coupled receptors (GPCRs), LDL receptors, cerberus and cerberus-like proteins, and proteins involved in carbohydrate metabolism, particularly those involved in epithelial development. Identifying the roles of these genes and their expression products may permit the definition of disease pathways and the identification of diagnostically and therapeutically useful targets.

GPCRs have been characterized as having seven putative transmembrane domains (designated TM1, TM2, TM3, TM4, TM5, TM6, and TM7) which are believed to represent transmembrane α-helices connected by extracellular or cytoplasmic loops. Most G-protein coupled receptors have single conserved cysteine residues in each of the first two extracellular loops which form disulfide bonds that are believed to stabilize functional protein structure. G-protein coupled receptors can be intracellularly coupled by heterotrimeric G-proteins to various intracellular enzymes, ion channels and transporters. Different G-protein α-subunits preferentially stimulate particular effectors to modulate various biological functions in a cell.

Platelet-activating factor (PAF) has been implicated as a mediator in diverse pathologic processes, such as allergy, asthma, septic shock, arterial thrombosis, and inflammatory processes (Prescott *et al., J. Biol. Chem.* 265:17381-17384 (1990)). PAF is a phospholipid (1-0-alkyl-2-acetyl-sn-glycero-3-phosphorylcholine) and exerts its various effects via specific cell surface receptors that bind PAF with high affinity. Using a guinea pig probe, Seyfried *et al.* (*Genomics* 13: 832-834 (1992)) isolated the gene for human PAF receptor (PTAFR). The coding sequence contains no introns. The encoded protein is highly homologous to the guinea pig PAF receptor (82% identity) and contains 7

-2-

putative transmembrane domains. The PAF receptor (PAFR) therefore appears to be a member of the G protein-coupled family of receptors and exhibits significant similarity to many members of this family. By analysis of rodent/human somatic cell hybrids, Seyfried et al. concluded that the PTAFR gene is located on human chromosome 1.

5   Recently, murine ESTs were identified (accession AA170490; GI:1749042; accession AA162789; GI:1738456) bearing sequence similarity to human platelet activating receptor homolog (accession NM 013308; GI:7019400). Additionally, a murine EST was identified (accession AA274112; GI:1912557) bearing sequence similarity to a human GPCR (KIAA0001; Nomura et al., DNA Res 2(4):210 (1995)) and to human platelet-activating factor receptor. The start of the EST

10   corresponds to approximately position 85 of PAF receptor which is in the 1st extracellular loop. The strongest homology is to KIAA0001, which is a randomly identified GPCR sequence with similarity to anaphylatoxin C5a receptor and PAF receptor. Over the past 15 years, nearly 350 therapeutic agents targeting 7 transmembrane receptors have been successfully introduced onto the market. As these receptors have an established, proven history as therapeutic targets, a clear need exists for

15   identification and characterization of GPCRs which can play a role in preventing, ameliorating or correcting dysfunctions or diseases.

            Epidermal growth factor (EGF) has a characteristic structure, which comprises three disulfide bridges that are essential for the protein's growth-stimulating activity. Many other proteins, however, have similar EGF domains (EGF-like domains). These proteins include both growth factors and other

20   proteins having functions unrelated to EGF. (See, e.g., Van Zoelen et al., Vitamins and Hormones 59:99-131 (2000)).

            Calcium binding EGF-like domains are present in many extracellular proteins that perform a diverse range of functions. These proteins include, for example, fibrillin-1, Notch-3, protein S, factor IX, and the low-density lipoprotein (LDL) receptor. (See, e.g., Smallridge et al., Journal of

25   Molecular Biology 286(3):661-668 (1999)).

            The LDL receptor family is known to bind and internalize apoE-rich lipoproteins and thought to play a role in lipoprotein metabolism. (See, e.g., Bujo, Nippon rinsho 57(12):2690-2695 (1999); Gliemann, Biological Chemistry 379:951-964 (1998)). Mutations cause familial hypercholesterolemia and premature coronary artery disease. LDL receptor-related proteins play an

30   important role in the clearance of plasma-activated alpha 2-acroglobulin and apolipoprotein E-enriched lipoproteins. It is essential for fetal development and has been associated with Alzheimer's disease (Hussain, et al., Ann. Rev. Nutr.19:141-72 (1999)). Recent studies have shown that this receptor family is important in neural development, suggesting that the LDL receptor family comprises multifunctional receptors involved in intracellular signal transduction, neuron migration,

35   vascular smooth muscle cell proliferation, and lipoprotein incorporation. (See, e.g., Bujo (1999), supra).

-3-

Recently, an expressed sequence tag (EST) (GenBank Accession No. AA482431) having EGF-like domains and showing similarities to LDL receptor-related proteins was identified. The gene, termed low-density lipoprotein receptor related protein 5 (LRP5), has been implicated as a candidate for conferring susceptibility to diabetes based on its chromosomal localization within the insulin-dependent diabetes mellitus (IDDM) locus (*see* Hey, *et al.*, *Gene* 216:103-111 (1998)).

Cerberus and cerberus-like proteins comprise a protein family expressed in the embryo, that are able to induce, enhance, and/or inhibit the formation, growth, and maintenance of neurons and related neural cells. The cerberus gene encodes a secreted protein that is expressed in the anterior endomesoderm, and when microinjected into Xenopus embryos, induces ectopic heads, and dupli-cated hearts and livers. (*See, e.g.*, Bouwmeester *et al.*, *Nature* 382(6592):595-601 (1996)). The cerberus-like (cerr1) gene was subsequently isolated in mice, and encodes a novel secreted protein specifically expressed in the anterior visceral endoderm during early gastrulation. (*See, e.g.*, Belo *et al.*, *Mechanisms of Development* 68:45-57 (1997)). In Xenopus assays, cer-1 acts as a potent neutralizing factor that induces forebrain markers and endoderm, but is unable to induce ectopic head-like structures as cerberus. Cer-1 encodes a putative secreted protein that is 48% identical to cerberus over a 110-amino acid region. (*See, e.g.*, Shawlot *et al.*, *Proceedings of the National Academy of Sciences, USA* 95(11):6198-6203 (1998)). These proteins appear to have a 9 cysteine residue pattern. (*See, e.g.*, U.S. Patent No. 5,935,852).

The cerberus protein has also shown to be a multifunctional antagonist of Nodal, BMP, and Wnt signals. Specifically, the cerberus protein functions as a multivalent growth factor antagonist in the extracellular space, binding to Nodal, BMP, and Wnt proteins through independent sites. (*See, e.g.*, Piccolo *et al.*, *Nature* 382 (6592):595-601 (1996)). Reported studies indicated that cerberus does not have a receptor or a dedicated transduction pathway, but acts as an extracellular inhibitor. (*See, e.g.*, Agius *et al.*, *Journal de la Societe de biologie* 193:347-354 (1999)).

The human cerberus protein has been isolated and reported. (*See, e.g.*, PCT Patent Publication No.: *WO 9849296*). As described in the art, these proteins are capable of inducing endodermal, cardiac, and neural tissue development in vertebrates when expressed, and may be useful in applications that require regeneration, differentiation, or repair of tissues, such as wound repair, neuronal regeneration or transplantation, supplementation of heart muscle differentiation, and differentiation of the pancreas. (*See, e.g.*, PCT Patent Publication No.: *WO 9748275*). These proteins may also be useful in the treatment of BMP-related disorders. (*See, e.g.*, *WO 9849296*). Recently, an expressed sequence tag (EST) (GenBank Accession No. AA120122) bearing high similarities to cerberus and cerberus-like genes was identified.

Successful development of fertile eggs involves the direct interaction between germline cells and follicle cells. (*See, e.g.*, McLaren and Wylie, *Current Problems in Germ Cell Differentiation*, Cambridge University Press (1992)). Epithelial morphogenesis and the development of the follicular

-4-

epithelium and germline differentiation that occurs in oogenesis have been studied well in *Drosophila melanogaster*. (*See, e.g.*, King, *Ovarian Development In Drosophila Melanogaster*, New York, Academic Press, (1970)). More particularly, the monolayer follicular epithelium has developmental, morphological, and molecular properties of vertebrate epithelia, and develops in concert with the

5   differentiating germline during oogenesis in *Drosophila Melanogaster*. (*See, e.g.*, Goode *et al.*, *Development* 122:3863-3879 (1996)).

Several classes of cell adhesion molecules are essential for epithelial development in both invertebrates and vertebrates. (*See, e.g.*, Gumbiner *et al.*, *Cell* 69:385-387 (1992)). For example, e-cadhedrin cell adhesion molecules are required for the formation and maintenance of epithelial

10   structure in vertebrates and *Drosophila*. (*See, e.g.*, Takeichie, *Development* 102:639-655 (1995); Tepass *et al.*, *Genes Dev.* 10:672-685 (1996); and Uemural *et al.*, *Genes Dev.* 10:659-671 (1996)). Receptor tyrosine kinases and their ligands are cell surface factors that are also necessary for epithelial development. (*See, e.g.*, Naldini *et al.*, *EMBO J.* 10:2867-2878 (1991); Miettinen *et al.*, *Nature* 376:337-341 (1995); Luetteke *et al.*, *Cell* 73:249-261 (1993); Schübach, *Cell* 49:699-707

15   (1987); Goode *et al.*, *Dev. Biol.* 178:35-50 (1996)).

Neurogenic genes encode another class of molecules important in the development of the epithelium. *Drosophila* neurogenic genes, including, Notch (n), Delta, neuralized, and Enhancer split are necessary for development of epithelial characteristics in embryonic tissue. (*See, e.g.*, Artavanis-Tsakonas *et al.*, *Science* 268:225-232 (1995); Hartenstein *et al.*, *Development* 116:1203-1220 (1992);

20   Coffman *et al.*, *Cell* 73:659-671 (1993); Ruohola *et al.*, *Cell* 66:1-20 (1991); and Xu *et al.*, *Development* 115:913-922 (1992)). Although these *Drosophila* neurogenic genes contribute to epithelial characteristics in embryonic tissue, little is known about their role in follicular epithelium morphogenesis.

One neurogenic gene, in particular, known as brainiac has shown to play a specific role in

25   epithelial development. The brainiac gene has been reported to encode a novel, putative secreted protein required in the germline for establishing the follicular epithelium and for determining its dorsal-ventral polarity. (*See, e.g.*, Goode *et al.*, (1992) and Goode *et al.*, *Dev. Biol.* (1996) *supra*). More specifically, the brainiac gene is present on the X chromosome and encodes a 325 amino acid protein with a putative signal sequence. The brainiac gene is expressed constitutively in the germline

30   during the first 12 hours of embryogenesis. (*See, e.g.*, Goode, S. *et al.* (1992) *supra*; Goode *et al.*, *Dev. Biol.* (1996) *supra*; and Goode *et al.*, *Development* (1996) *supra*) Mutations in the brainiac gene demonstrate synergistic genetic interactions with mutations in signaling molecules, including, transforming growth factor alpha (TGF-α) and epidermal growth factor receptor (Egfr). TGF-α and Egfr mutant mice suffer from multi-organ failure due to widespread impairment of epithelial

35   development (*See, e.g.*, Miettinen *et al.*, (1995) *supra*).

-5-

Brainiac genes are important for correct development of the follicular epithelium. Brainiac mutant females and their offspring have multiple defects including ventralization of the eggshell, gaps in the follicular epithelium, and multiple layers of follicle cells around oocytes. In the absence of brainiac gene function, less efficiency in follicular epithelium formation occurs, resulting in a failure

5   to individuate germline cysts. Once the follicular epithelium is formed, brainiac genes are essential for maintenance of its epithelial characteristics. In addition, disruption in brainiac germline function results in loss of apicalbasal polarity and an accumulation of multiple layers of follicle cells, particularly around the oocyte. Brainiac is essential for efficient migration and maintenance of border cells and main body epithelial cells moving towards the oocyte during the late phases of oogenesis.

10   (*See, e.g.*, Goode *et al.*, *Development* (1996) *supra*). An additional gene known as egghead has shown to function similarly in epithelial development. Moreover, both genes appear to be components of novel signaling pathways essential for epithelial development and maintenance. Most recently, an expressed sequence tag (EST) was isolated bearing sequence similarity to genes encoding neurogenic secreted signaling proteins, and more particularly to the brainiac gene (GenBank

15   Accession No.: AA133340; EST name: zn92h01.s1).

Summary of the Invention

The present invention generally relates to transgenic animals, as well as to compositions and methods relating to the characterization of gene function.

The present invention provides transgenic cells comprising a disruption in a targeted gene.

20   The transgenic cells of the present invention are comprised of any cells capable of undergoing homologous recombination. Preferably, the cells of the present invention are stem cells and more preferably, embryonic stem (ES) cells, and most preferably, murine ES cells. According to one embodiment, the transgenic cells are produced by introducing a targeting construct into a stem cell to produce a homologous recombinant, resulting in a mutation of the targeted gene. In another

25   embodiment, the transgenic cells are derived from the transgenic animals described below. The cells derived from the transgenic animals includes cells that are isolated or present in a tissue or organ, and any cell lines or any progeny thereof.

The present invention also provides a targeting construct and methods of producing the targeting construct that when introduced into stem cells produces a homologous recombinant. In one

30   embodiment, the targeting construct of the present invention comprises first and second polynucleotide sequences that are homologous to the targeted gene. The targeting construct also comprises a polynucleotide sequence that encodes a selectable marker that is preferably positioned between the two different homologous polynucleotide sequences in the construct. The targeting construct may also comprise other regulatory elements that may enhance homologous recombination.

35   The present invention further provides non-human transgenic animals and methods of producing such non-human transgenic animals comprising a disruption in a target gene. The

-6-

transgenic animals of the present invention include transgenic animals that are heterozygous and homozygous for a mutation in the target gene. In one aspect, the transgenic animals of the present invention are defective in the function of the target gene. In another aspect, the transgenic animals of the present invention comprise a phenotype associated with having a mutation in a target gene.

5   The present invention also provides methods of identifying agents capable of affecting a phenotype associated with a disruption of a target gene in a transgenic animal. For example, a putative agent is administered to the transgenic animal and a response of the transgenic animal to the putative agent is measured and compared to the response of a "normal" or wild type mouse, or alternatively compared to a transgenic animal control (without agent administration). The invention

10   further provides agents identified according to such methods. The present invention also provides methods of identifying agents useful as therapeutic agents for treating conditions associated with a disruption of the target gene.

The present invention further provides a method of identifying agents having an effect on target gene expression or function. The method includes administering an effective amount of the

15   agent to a transgenic animal, preferably a mouse. The method includes measuring a response of the transgenic animal, for example, to the agent, and comparing the response of the transgenic animal to a control animal, which may be, for example, a wild-type animal or alternatively, a transgenic animal control. Compounds that may have an effect on target gene expression or function may also be screened against cells in cell-based assays, for example, to identify such compounds.

20   The invention also provides cell lines comprising nucleic acid sequences of a target gene. Such cell lines may be capable of expressing such sequences by virtue of operable linkage to a promoter functional in the cell line. Preferably, expression of the target gene sequence is under the control of an inducible promoter. Also provided are methods of identifying agents that interact with the target gene, comprising the steps of contacting the target gene or target protein with an agent and

25   detecting an agent/ target gene or agent/target protein complex. Such complexes can be detected by, for example, measuring expression of an operably linked detectable marker.

The invention further provides methods of treating diseases or conditions associated with a disruption in a target gene, and more particularly, to a disruption in the expression or function of the target gene or the target protein encoded by the target gene. In a preferred embodiment, methods of

30   the present invention involve treating diseases or conditions associated with a disruption in the target gene's expression or function, including administering to a subject in need, a therapeutic agent which effects target gene or target protein expression or function. In accordance with this embodiment, the method comprises administration of a therapeutically effective amount of a natural, synthetic, semi-synthetic, or recombinant target gene, target gene products or fragments thereof as well as natural,

35   synthetic, semi-synthetic or recombinant analogs.

-7-

The present invention further provides methods of treating diseases or conditions associated with disrupted targeted gene expression or function, wherein the methods comprise detecting and replacing through gene therapy mutated target genes.

Definitions

5   The term "gene" refers to (a) a gene containing at least one of the DNA sequences disclosed herein; (b) any DNA sequence that encodes the amino acid sequence encoded by the DNA sequences disclosed herein and/or; (c) any DNA sequence that hybridizes to the complement of the coding sequences disclosed herein. Preferably, the term includes coding as well as noncoding regions, and preferably includes all sequences necessary for normal gene expression including promoters,

10   enhancers and other regulatory sequences.

The terms "polynucleotide" and "nucleic acid molecule" are used interchangeably to refer to polymeric forms of nucleotides of any length. The polynucleotides may contain deoxyribonucleo- tides, ribonucleotides and/or their analogs. Nucleotides may have any three-dimensional structure, and may perform any function, known or unknown. The term "polynucleotide" includes single-,

15   double-stranded and triple helical molecules.

"Oligonucleotide" refers to polynucleotides of between 5 and about 100 nucleotides of single- or double-stranded DNA. Oligonucleotides are also known as oligomers or oligos and may be isolated from genes, or chemically synthesized by methods known in the art. A "primer" refers to an oligonucleotide, usually single-stranded, that provides a 3'-hydroxyl end for the initiation of enzyme-

20   mediated nucleic acid synthesis. The following are non-limiting embodiments of polynucleotides: a gene or gene fragment, exons, introns, mRNA, tRNA, rRNA, ribozymes, cDNA, recombinant polynucleotides, branched polynucleotides, plasmids, vectors, isolated DNA of any sequence, isolated RNA of any sequence, nucleic acid probes and primers. A nucleic acid molecule may also comprise modified nucleic acid molecules, such as methylated nucleic acid molecules and nucleic acid

25   molecule analogs. Analogs of purines and pyrimidines are known in the art, and include, but are not limited to, aziridinycytosine, 4-acetylcytosine, 5-fluorouracil, 5-bromouracil, 5-carboxymethylamino- methyl-2-thiouracil, 5-carboxymethyl-aminomethyluracil, inosine, N6-isopentenyladenine, 1-methyl- adenine, 1-methylpseudouracil, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyl- adenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, pseudouracil, 5-pentylnyluracil and

30   2,6-diaminopurine. The use of uracil as a substitute for thymine in a deoxyribonucleic acid is also considered an analogous form of pyrimidine.

A "fragment" of a polynucleotide is a polynucleotide comprised of at least 9 contiguous nucleotides, preferably at least 15 contiguous nucleotides and more preferably at least 45 nucleotides, of coding or non-coding sequences.

-8-

The term "gene targeting" refers to a type of homologous recombination that occurs when a fragment of genomic DNA is introduced into a mammalian cell and that fragment locates and recombines with endogenous homologous sequences.

The term "homologous recombination" refers to the exchange of DNA fragments between two DNA molecules or chromatids at the site of homologous nucleotide sequences.

The term "homologous" as used herein denotes a characteristic of a DNA sequence having at least about 70 percent sequence identity as compared to a reference sequence, typically at least about 85 percent sequence identity, preferably at least about 95 percent sequence identity, and more preferably about 98 percent sequence identity, and most preferably about 100 percent sequence identity as compared to a reference sequence. Homology can be determined using a "BLASTN" algorithm. It is understood that homologous sequences can accommodate insertions, deletions and substitutions in the nucleotide sequence. Thus, linear sequences of nucleotides can be essentially identical even if some of the nucleotide residues do not precisely correspond or align. The reference sequence may be a subset of a larger sequence, such as a portion of a gene or flanking sequence, or a repetitive portion of a chromosome.

The term "target gene" (alternatively referred to as "target gene sequence" or "target DNA sequence" or "target sequence") refers to any nucleic acid molecule or polynucleotide of any gene to be modified by homologous recombination. The target sequence includes an intact gene, an exon or intron, a regulatory sequence or any region between genes. The target gene comprises a portion of a particular gene or genetic locus in the individual's genomic DNA. As provided herein, the target gene of the present invention is a platelet activating receptor gene, a PAF receptor gene, an LRP5 gene, a cerberus gene, and a brainiac gene.

A "platelet activating receptor gene" refers to a sequence comprising SEQ ID NO:1 or comprising the sequence identified in Genebank as Accession No. AA162789; GI:1738456. In one aspect, the coding sequence of the platelet activating receptor gene comprises SEQ ID NO:1 or comprises the gene identified in Genebank as Accession No.: AA162789; GI:1738456.

An "PAF receptor gene" refers to a sequence comprising SEQ ID NO:4 or comprising the sequence identified in Genebank as Accession No. AA274112; GI:1912557. In one aspect, the coding sequence of the PAF receptor gene comprises SEQ ID NO:4 or comprises the gene identified in Genebank as Accession No. AA274112; GI:1912557.

An "LPR5 gene" refers to a sequence comprising refers to a sequence comprising SEQ ID NO:7 or comprising the sequence identified in Genebank as Accession No. NM_008513; GI:6678715. In one aspect, the coding sequence of the LPR5 gene comprises SEQ ID NO:7 or comprises the gene identified in Genebank as Accession No. NM_008513; GI:6678715.

A "cerberus gene" refers to a sequence comprising SEQ ID NO:10 or comprising the sequence identified in Genebank as Accession No. NM_009887; GI:6753409. In one aspect, the

-9-

coding sequence of the cerberus gene comprises SEQ ID NO:7 or comprises the gene identified in Genebank as Accession No. NM_009887; GI:6753409.

A "brainiac gene" refers to a sequence comprising SEQ ID NO:13 or comprising the sequence identified in GenBank as Accession No.: AA133340; EST name: zn92h01.s1. In one aspect, the

5   coding sequence of the platelet activating receptor gene comprises SEQ ID NO:13 or comprises the gene identified in Genebank as Accession No.: AA133340; EST name: zn92h01.s1.

"Disruption" of a target gene occurs when a fragment of genomic DNA locates and recombines with an endogenous homologous sequence. These sequence disruptions or modifications may include insertions, missense, frameshift, deletion, or substitutions, or replacements of DNA

10   sequence, or any combination thereof. Insertions include the insertion of entire genes which may be of animal, plant, prokaryotic, or viral origin. Disruption, for example, can alter or replace a promoter, enhancer, or splice site of a target gene, and can alter the normal gene product by inhibiting its production partially or completely or by enhancing the normal gene product's activity.

The term, "transgenic cell", refers to a cell containing within its genome a target gene that has

15   been disrupted, modified, altered, or replaced completely or partially by the method of gene targeting.

The term "transgenic animal" refers to an animal that contains within its genome a specific gene that has been disrupted by the method of gene targeting. The transgenic animal includes both the heterozygote animal (*i.e.*, one defective allele and one wild-type allele) and the homozygous animal (*i.e.*, two defective alleles).

20   As used herein, the terms "selectable marker" or "positive selection marker" refers to a gene encoding a product that enables only the cells that carry the gene to survive and/or grow under certain conditions. For example, plant and animal cells that express the introduced neomycin resistance (Neo^r) gene are resistant to the compound G418. Cells that do not carry the Neo^r gene marker are killed by G418. Other positive selection markers will be known to those of skill in the art.

25   A "host cell" includes an individual cell or cell culture which can be or has been a recipient for vector(s) or for incorporation of nucleic acid molecules and/or proteins. Host cells include progeny of a single host cell, and the progeny may not necessarily be completely identical (in morphology or in total DNA complement) to the original parent due to natural, accidental, or deliberate mutation. A host cell includes cells transfected with the constructs of the present invention.

30   The term "modulates" as used herein refers to the inhibition, reduction, increase or enhancement of a target gene or target protein function, expression, activity, or alternatively a phenotype associated with a disruption in a target gene.

The term "ameliorates" refers to a decreasing, reducing or eliminating of a condition, disease, disorder, or phenotype, including an abnormality or symptom associated with a disruption in a target

35   gene.

-10-

The term "abnormality" refers to any disease, disorder, condition, or phenotype in which a disruption of a target gene is implicated, including pathological conditions.

Brief Description of the Drawings

Figure 1 shows the polynucleotide sequence for a platelet activating receptor gene (SEQ ID

5   NO:1).

Figure 2A-2B shows design of the targeting construct used to disrupt platelet activating receptor genes. Figure 2B shows the sequences identified as SEQ ID NO:2 and SEQ ID NO:3, which were used as the targeting arms (homologous sequences) in the targeting construct.

Figure 3 shows a graph relating to the hot plate testing on mice having a disruption in a

10   platelet activating receptor genes.

Figure 4 shows a graph relating to the open field testing on mice having a disruption in a platelet activating receptor genes.

Figure 5 shows the polynucleotide sequence for a PAF receptor gene (SEQ ID NO:4).

Figure 6A-6B shows design of the targeting construct used to disrupt PAF genes. Figure 6B

15   shows the sequences identified as SEQ ID NO:5 and SEQ ID NO:6, which were used as the targeting arms (homologous sequences) in the targeting construct.

Figure 7A-7C shows the polynucleotide sequence for a LPR5 gene (SEQ ID NO:7).

Figure 8 shows the amino acid sequence for a LPR5 polypeptide (SEQ ID NO:8).

Figure 9A-9C shows design of the targeting construct used to disrupt LPR5 genes. Figure 9C

20   shows the sequences identified as SEQ ID NO:9 and SEQ ID NO:10, which were used as the targeting arms (homologous sequences) in the targeting construct.

Figure 10 shows a graph relating to the open field testing (time spent in the central region) of mice having a disruption in a LPR5 gene.

Figure 11 shows a graph relating to the open field testing (total distance traveled) of mice

25   having a disruption in a LPR5 gene.

Figure 12 shows the polynucleotide sequence for the cerberus gene (SEQ ID NO:11) and the amino acid sequence of a cerberus polypeptide (SEQ ID NO:12).

Figure 13A-13B shows design of the targeting construct used to disrupt cerberus genes. Figure 13B shows the sequences identified as SEQ ID NO:13 and SEQ ID NO:14, which were used

30   as the targeting arms (homologous sequences) in the targeting construct.

Figure 14 shows a graph relating to the open field testing of mice having a disruption in a cerberus gene.

Figure 15 shows a graph relating to the tail suspension testing of mice having a disruption in a cerberus gene.

-11-

Figure 16 shows the polynucleotide sequence of a brainiac gene (SEQ ID NO:17) and SEQ ID NO:16 and SEQ ID NO:17 which were used as the targeting arms (homologous sequences) in the brainiac gene targeting construct.

Detailed Description of the Invention

5   The invention is based, in part, on the evaluation of the expression and role of genes and gene expression products, primarily those associated with a target gene. Among others, the invention permits the definition of disease pathways and the identification of diagnostically and therapeutically useful targets. For example, genes which are mutated or down-regulated under disease conditions may be involved in causing or exacerbating the disease condition. Treatments directed at up-

10   regulating the activity of such genes or treatments which involve alternate pathways, may ameliorate the disease condition.

Generation of Targeting Construct

The targeting construct of the present invention may be produced using standard methods known in the art. (See, e.g., Sambrook, et al., 1989, Molecular Cloning: A Laboratory Manual,

15   Second Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York; E.N. Glover (eds.), 1985, DNA Cloning: A Practical Approach, Volumes I and II; M.J. Gait (ed.), 1984, Oligonucleotide Synthesis; B.D. Hames & S.J. Higgins (eds.), 1985, Nucleic Acid Hybridization; B.D. Hames & S.J. Higgins (eds.), 1984, Transcription and Translation; R.I. Freshney (ed.), 1986, Animal Cell Culture; Immobilized Cells and Enzymes, IRL Press, 1986; B. Perbal, 1984, A Practical Guide

20   To Molecular Cloning; F.M. Ausubel et al., 1994, Current Protocols in Molecular Biology, John Wiley & Sons, Inc.). For example, the targeting construct may be prepared in accordance with conventional ways, where sequences may be synthesized, isolated from natural sources, manipulated, cloned, ligated, subjected to in vitro mutagenesis, primer repair, or the like. At various stages, the joined sequences may be cloned, and analyzed by restriction analysis, sequencing, or the like.

25   The targeting DNA can be constructed using techniques well known in the art. For example, the targeting DNA may be produced by chemical synthesis of oligonucleotides, nick-translation of a double-stranded DNA template, polymerase chain-reaction amplification of a sequence (or ligase chain reaction amplification), purification of prokaryotic or target cloning vectors harboring a sequence of interest (e.g., a cloned cDNA or genomic DNA, synthetic DNA or from any of the

30   aforementioned combination) such as plasmids, phagemids, YACs, cosmids, bacteriophage DNA, other viral DNA or replication intermediates, or purified restriction fragments thereof, as well as other sources of single and double-stranded polynucleotides having a desired nucleotide sequence. Moreover, the length of homology may be selected using known methods in the art. For example, selection may be based on the sequence composition and complexity of the predetermined

35   endogenous target DNA sequence(s).

-12-

The targeting construct of the present invention typically comprises a first sequence homologous to a portion or region of the target gene and a second sequence homologous to a second portion or region of the target gene. The targeting construct further comprises a positive selection marker, which is preferably positioned in between the first and the second DNA sequence that are

5   homologous to a portion or region of the target DNA sequence. The positive selection marker may be operatively linked to a promoter and a polyadenylation signal.

Other regulatory sequences known in the art may be incorporated into the targeting construct to disrupt or control expression of a particular gene in a specific cell type. In addition, the targeting construct may also include a sequence coding for a screening marker, for example, green fluorescent

10   protein (GFP), or another modified fluorescent protein.

Although the size of the homologous sequence is not critical and can range from as few as 50 base pairs to as many as 100 kb, preferably each fragment is greater than about 1 kb in length, more preferably between about 1 and about 10 kb, and even more preferably between about 1 and about 5 kb. One of skill in the art will recognize that although larger fragments may increase the number of

15   homologous recombination events in ES cells, larger fragments will also be more difficult to clone.

In a preferred embodiment of the present invention, the targeting construct is prepared directly from a plasmid genomic library using the methods described in pending U.S. Patent Application No.: 08/971,310, filed November 17, 1997, the disclosure of which is incorporated herein in its entirety. Generally, a sequence of interest is identified and isolated from a plasmid library in a single

20   step using, for example, long-range PCR. Following isolation of this sequence, a second polynucleotide that will disrupt the target sequence can be readily inserted between two regions encoding the sequence of interest. In accordance with this aspect, the construct is generated in two steps by (1) amplifying (for example, using long-range PCR) sequences homologous to the target sequence, and (2) inserting another polynucleotide (for example a selectable marker) into the PCR product so

25   that it is flanked by the homologous sequences. Typically, the vector is a plasmid from a plasmid genomic library. The completed construct is also typically a circular plasmid.

In another embodiment, the targeting construct is designed in accordance with the regulated positive selection method described in U.S. Application No. 60/232,957, filed September 15, 2000, the disclosure of which is incorporated herein in its entirety. The targeting construct is designed to

30   include a PGK-neo fusion gene having two lacO sites, positioned in the PGK promoter and an NLS-lacI gene comprising a lac repressor fused to sequences encoding the NLS from the SV40 T antigen.

In another embodiment, the targeting construct may contain more than one selectable maker gene, including a negative selectable marker, such as the herpes simplex virus tk (HSV-tk) gene. The negative selectable marker may be operatively linked to a promoter and a polyadenylation signal.

35   (See, e.g., U.S. Patent No. 5,464,764; U.S. Patent No. 5,487,992; U.S. Patent No. 5,627,059; and U.S. Patent No. 5,631,153).

-13-

Generation of Cells and Confirmation of Homologous Recombination Events

Once an appropriate targeting construct has been prepared, the targeting construct may be introduced into an appropriate host cell using any method known in the art. Various techniques may be employed in the present invention, including, for example, pronuclear microinjection; retrovirus
5   mediated gene transfer into germ lines; gene targeting in embryonic stem cells; electroporation of embryos; sperm-mediated gene transfer; and calcium phosphate/DNA co-precipitates, microinjection of DNA into the nucleus, bacterial protoplast fusion with intact cells, transfection, polycations, *e.g.*, polybrene, polyornithine, *etc.*, or the like (*See, e.g.*, U.S. Pat. No. 4,873,191; Van der Putten, *et al.*, 1985, *Proc. Natl. Acad. Sci., USA* 82:6148-6152; Thompson, *et al.*, 1989, *Cell* 56:313-321; Lo, 1983,
10   *Mol Cell. Biol.* 3:1803-1814; Lavitrano, *et al.*,1989, *Cell*, 57:717-723). Various techniques for transforming mammalian cells are known in the art. (*See, e.g.*, Gordon, 1989, *Intl. Rev. Cytol.*, 115:171-229; Keown *et al.*, 1989, *Methods in Enzymology*; Keown *et al.*, 1990, *Methods and Enzymology*, Vol. 185, pp. 527-537; Mansour *et al.*, 1988, *Nature*, 336:348-352).

In a preferred aspect of the present invention, the targeting construct is introduced into host
15   cells by electroporation. In this process, electrical impulses of high field strength reversibly permeabilize biomembranes allowing the introduction of the construct. The pores created during electroporation permit the uptake of macromolecules such as DNA. (*See, e.g.*, Potter, H., *et al.*, 1984, *Proc. Nat'l. Acad. Sci. U.S.A.* 81:7161-7165).

Any cell type capable of homologous recombination may be used in the practice of the
20   present invention. Examples of such target cells include cells derived from vertebrates including mammals such as humans, bovine species, ovine species, murine species, simian species, and ether eucaryotic organisms such as filamentous fungi, and higher multicellular organisms such as plants.

Preferred cell types include embryonic stem (ES) cells, which are typically obtained from pre-implantation embryos cultured *in vitro*. (*See, e.g.*, Evans, M. J., *et al.*, 1981, *Nature* 292:154-156;
25   Bradley, M. O., *et al.*, 1984, *Nature* 309:255-258; Gossler *et al.*, 1986, *Proc. Natl. Acad. Sci. USA* 83:9065-9069; and Robertson, *et al.*, 1986, *Nature* 322:445-448). The ES cells are cultured and prepared for introduction of the targeting construct using methods well known to the skilled artisan. (*See, e.g.*, Robertson, E. J. ed. "Teratocarcinomas and Embryonic Stem Cells, a Practical Approach", IRL Press, Washington D.C., 1987; Bradley *et al.*, 1986, *Current Topics in Devel. Biol.* 20:357-371;
30   by Hogan *et al.* in "Manipulating the Mouse Embryo": A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor N.Y., 1986; Thomas *et al.*, 1987, *Cell* 51:503; Koller *et al.*, 1991, *Proc. Natl. Acad. Sci. USA*, 88:10730; Dorin *et al.*, 1992, *Transgenic Res.* 1:101; and Veis *et al.*, 1993, *Cell* 75:229). The ES cells that will be inserted with the targeting construct are derived from an embryo or blastocyst of the same species as the developing embryo into which they are to be
35   introduced. ES cells are typically selected for their ability to integrate into the inner cell mass and contribute to the germ line of an individual when introduced into the mammal in an embryo at the

-14-

blastocyst stage of development. Thus, any ES cell line having this capability is suitable for use in the practice of the present invention.

5   The present invention may also be used to knockout genes in other cell types, such as stem cells. By way of example, stem cells may be myeloid, lymphoid, or neural progenitor and precursor cells. These cells comprising a disruption or knockout of a gene may be particularly useful in the study of target gene function in individual developmental pathways. Stem cells may be derived from any vertebrate species, such as mouse, rat, dog, cat, pig, rabbit, human, non-human primates and the like.

10   After the targeting construct has been introduced into cells, the cells where successful gene targeting has occurred are identified. Insertion of the targeting construct into the targeted gene is typically detected by identifying cells for expression of the marker gene. In a preferred embodiment, the cells transformed with the targeting construct of the present invention are subjected to treatment with an appropriate agent that selects against cells not expressing the selectable marker. Only those cells expressing the selectable marker gene survive and/or grow under certain conditions. For

15   .   example, cells that express the introduced neomycin resistance gene are resistant to the compound G418, while cells that do not express the neo gene marker are killed by G418. If the targeting construct also comprises a screening marker such as GFP, homologous recombination can be identified through screening cell colonies under a fluorescent light. Cells that have undergone homologous recombination will have deleted the GFP gene and will not fluoresce.

20   If a regulated positive selection method is used in identifying homologous recombination events, the targeting construct is designed so that the expression of the selectable marker gene is regulated in a manner such that expression is inhibited following random integration but is permitted (derepressed) following homologous recombination. More particularly, the transfected cells are screened for expression of the *neo* gene, which requires that (1) the cell was successfully

25   electroporated, and (2) *lac* repressor inhibition of *neo* transcription was relieved by homologous recombination. This method allows for the identification of transfected cells and homologous recombinants to occur in one step with the addition of a single drug.

Alternatively, a positive-negative selection technique may be used to select homologous recombinants. This technique involves a process in which a first drug is added to the cell population,

30   for example, a neomycin-like drug to select for growth of transfected cells, *i.e.* positive selection. A second drug, such as FIAU is subsequently added to kill cells that express the negative selection marker, *i.e.* negative selection. Cells that contain and express the negative selection marker are killed by a selecting agent, whereas cells that do not contain and express the negative selection marker survive. For example, cells with non-homologous insertion of the construct express HSV thymidine

35   kinase and therefore are sensitive to the herpes drugs such as gancyclovir (GANC) or FIAU (1-(2-

-15-

deoxy 2-fluoro-B-D-arabinofluranosyl)-5-iodouracil). (*See, e.g.*, Mansour *et al., Nature* 336:348-352:
(1988); Capecchi, *Science* 244:1288-1292, (1989); Capecchi, *Trends in Genet.* 5:70-76 (1989)).

Successful recombination may be identified by analyzing the DNA of the selected cells to
confirm homologous recombination. Various techniques known in the art, such as PCR and/or
Southern analysis may be used to confirm homologous recombination events.

Homologous recombination may also be used to disrupt genes in stem cells, and other cell
types, which are not totipotent embryonic stem cells. By way of example, stem cells may be myeloid,
lymphoid, or neural progenitor and precursor cells. Such transgenic cells may be particularly useful
in the study of gene function in individual developmental pathways. Stem cells may be derived from
any vertebrate species, such as mouse, rat, dog, cat, pig, rabbit, human, non-human primates and the
like.

In cells which are not totipotent it may be desirable to knock out both copies of the target
using methods which are known in the art. For example, cells comprising homologous recombination
at a target locus which have been selected for expression of a positive selection marker (e.g., Neo^r)
and screened for non-random integration, can be further selected for multiple copies of the selectable
marker gene by exposure to elevated levels of the selective agent (e.g., G418). The cells are then
analyzed for homozygosity at the target locus. Alternatively, a second construct can be generated
with a different positive selection marker inserted between the two homologous sequences. The two
constructs can be introduced into the cell either sequentially or simultaneously, followed by
appropriate selection for each of the positive marker genes. The final cell is screened for homologous
recombination of both alleles of the target.

Production of Transgenic Animals

Selected cells are then injected into a blastocyst (or other stage of development suitable for
the purposes of creating a viable animal, such as, for example, a morula) of an animal (*e.g.*, a mouse)
to form chimeras (*see e.g.*, Bradley, A. *in Teratocarcinomas and Embryonic Stem Cells: A Practical
Approach*, E. J. Robertson, ed., IRL, Oxford, pp. 113-152 (1987)). Alternatively, selected ES cells
can be allowed to aggregate with dissociated mouse embryo cells to form the aggregation chimera. A
chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the
embryo brought to term. Chimeric progeny harboring the homologously recombined DNA in their
germ cells can be used to breed animals in which all cells of the animal contain the homologously
recombined DNA. In one embodiment, chimeric progeny mice are used to generate a mouse with a
heterozygous disruption in the gene. Heterozygous transgenic mice can then be mated. It is well
know in the art that typically ¼ of the offspring of such matings will have a homozygous disruption in
the target gene.

The heterozygous and homozygous transgenic mice can then be compared to normal, wild
type mice to determine whether disruption of the target gene causes phenotypic changes, especially

-16-

pathological changes. For example, heterozygous and homozygous mice may be evaluated for

phenotypic changes by physical examination, necropsy, histology, clinical chemistry, complete blood

count, body weight, organ weights, and cytological evaluation of bone marrow.

5   In one embodiment, the phenotype (or phenotypic change) associated with a disruption in the

target gene is placed into or stored in a database. Preferably, the database includes: (i) genotypic data

(e.g., identification of the disrupted gene) and (ii) phenotypic data (e.g., phenotype(s) resulting from

the gene disruption) associated with the genotypic data. The database is preferably electronic. In

addition, the database is preferably combined with a search tool so that the database is searchable.

Conditional Transgenic Animals

10   The present invention further contemplates conditional transgenic or knockout animals, such

as those produced using recombination methods. Bacteriophage P1 Cre recombinase and flp

recombinase from yeast plasmids are two non-limiting examples of site-specific DNA recombinase

enzymes which cleave DNA at specific target sites (lox P sites for cre recombinase and frt sites for flp

recombinase) and catalyze a ligation of this DNA to a second cleaved site. A large number of suitable

15   alternative site-specific recombinases have been described, and their genes can be used in accordance

with the method of the present invention. Such recombinases include the Int recombinase of bacterio-

phage λ (with or without Xis) (Weisberg, R. et. al., in *Lambda II*, (Hendrix, R., *et al.*, Eds.), Cold

Spring Harbor Press, Cold Spring Harbor, NY, pp. 211-50 (1983), herein incorporated by reference);

TpnI and the β-lactamase transposons (Mercier, *et al., J. Bacteriol.*, 172:3745-57 (1990)); the Tn3

20   resolvase (Flanagan & Fennewald *J. Molec. Biol.*, 206:295-304 (1989); Stark, *et al., Cell*, 58:779-90

(1989)); the yeast recombinases (Matsuzaki, *et al., J. Bacteriol.*, 172:610-18 (1990)); the *B. subtilis*

SpoIVC recombinase (Sato, *et al., J. Bacteriol.* 172:1092-98 (1990)); the Flp recombinase (Schwartz

& Sadowski, *J. Molec.Biol.*, 205:647-658 (1989); Parsons, *et al., J. Biol. Chem.*, 265:4527-33 (1990);

Golic & Lindquist, *Cell*, 59:499-509 (1989); Amin, *et al., J. Molec. Biol.*, 214:55-72 (1990)); the Hin

25   recombinase (Glasgow, *et al., J. Biol. Chem.*, 264:10072-82 (1989)); immunoglobulin recombinases

(Malynn, *et al., Cell*, 54:453-460 (1988)); and the Cin recombinase (Haffter & Bickle, *EMBO J.*,

7:3991-3996 (1988); Hubner, *et al., J. Molec. Biol.*, 205:493-500 (1989)), all herein incorporated by

reference. Such systems are discussed by Echols (*J. Biol. Chem.* 265:14697-14700 (1990)); de

Villartay (*Nature*, 335:170-74 (1988)); Craig, (*Ann. Rev. Genet.*, 22:77-105 (1988)); Poyart-Salmeron,

30   *et al.*, (*EMBO J.* 8:2425-33 (1989)); Hunger-Bertling, *et al.* (*Mol Cell. Biochem.*, 92:107-16 (1990));

and Cregg & Madden (*Mol. Gen. Genet.*, 219:320-23 (1989)), all herein incorporated by reference.

Cre has been purified to homogeneity, and its reaction with the loxP site has been extensively

characterized (Abremski & Hess *J. Mol. Biol.* 259:1509-14 (1984), herein incorporated by reference).

Cre protein has a molecular weight of 35,000 and can be obtained commercially from New England

35   Nuclear/DuPont. The cre gene (which encodes the Cre protein) has been cloned and expressed

(Abremski, *et al. Cell* 32:1301-11 (1983), herein incorporated by reference). The Cre protein

-17-

mediates recombination between two loxP sequences (Sternberg, *et al. Cold Spring Harbor Symp. Quant. Biol.* 45:297-309 (1981)), which may be present on the same or different DNA molecule. Because the internal spacer sequence of the loxP site is asymmetrical, two loxP sites can exhibit directionality relative to one another (Hoess & Abremski *Proc. Natl. Acad. Sci. U.S.A.* 81:1026-29 (1984)). Thus, when two sites on the same DNA molecule are in a directly repeated orientation, Cre will excise the DNA between the sites (Abremski, *et al. Cell* 32:1301-11 (1983)). However, if the sites are inverted with respect to each other, the DNA between them is not excised after recombination but is simply inverted. Thus, a circular DNA molecule having two loxP sites in direct orientation will recombine to produce two smaller circles, whereas circular molecules having two loxP sites in an inverted orientation simply invert the DNA sequences flanked by the loxP sites. In addition, recombinase action can result in reciprocal exchange of regions distal to the target site when targets are present on separate DNA molecules.

Recombinases have important application for characterizing gene function in knockout models. When the constructs described herein are used to disrupt target genes, a fusion transcript can be produced when insertion of the positive selection marker occurs downstream (3') of the translation initiation site of the target gene. The fusion transcript could result in some level of protein expression with unknown consequence. It has been suggested that insertion of a positive selection marker gene can affect the expression of nearby genes. These effects may make it difficult to determine gene function after a knockout event since one could not discern whether a given phenotype is associated with the inactivation of a gene, or the transcription of nearby genes. Both potential problems are solved by exploiting recombinase activity. When the positive selection marker is flanked by recombinase sites in the same orientation, the addition of the corresponding recombinase will result in the removal of the positive selection marker. In this way, effects caused by the positive selection marker or expression of fusion transcripts are avoided.

In one embodiment, purified recombinase enzyme is provided to the cell by direct microinjection. In another embodiment, recombinase is expressed from a co-transfected construct or vector in which the recombinase gene is operably linked to a functional promoter. An additional aspect of this embodiment is the use of tissue-specific or inducible recombinase constructs which allow the choice of when and where recombination occurs. One method for practicing the inducible forms of recombinase-mediated recombination involves the use of vectors that use inducible or tissue-specific promoters or other gene regulatory elements to express the desired recombinase activity. The inducible expression elements are preferably operatively positioned to allow the inducible control or activation of expression of the desired recombinase activity. Examples of such inducible promoters or other gene regulatory elements include, but are not limited to, tetracycline, metallothionine, ecdysone, and other steroid-responsive promoters, rapamycin responsive promoters, and the like (No, *et al. Proc. Natl. Acad. Sci. USA*, 93:3346-51 (1996); Furth, *et al. Proc. Natl. Acad. Sci. USA*, 91:9302-6

-18-

(1994)). Additional control elements that can be used include promoters requiring specific transcription factors such as viral, promoters. Vectors incorporating such promoters would only express recombinase activity in cells that express the necessary transcription factors.

Models for Disease

The cell- and animal-based systems described herein can be utilized as models for diseases. Animals of any species, including, but not limited to, mice, rats, rabbits, guinea pigs, pigs, micro-pigs, goats, and non-human primates, e.g., baboons, monkeys, and chimpanzees may be used to generate disease animal models. In addition, cells from humans may be used. These systems may be used in a variety of applications. Such assays may be utilized as part of screening strategies designed to identify agents, such as compounds which are capable of ameliorating disease symptoms. Thus, the animal- and cell-based models may be used to identify drugs, pharmaceuticals, therapies and interventions which may be effective in treating disease.

Cell-based systems may be used to identify compounds which may act to ameliorate disease symptoms. For example, such cell systems may be exposed to a compound suspected of exhibiting an ability to ameliorate disease symptoms, at a sufficient concentration and for a time sufficient to elicit such an amelioration of disease symptoms in the exposed cells. After exposure, the cells are examined to determine whether one or more of the disease cellular phenotypes has been altered to resemble a more normal or more wild type, non-disease phenotype.

In addition, animal-based disease systems, such as those described herein, may be used to identify compounds capable of ameliorating disease symptoms. Such animal models may be used as test substrates for the identification of drugs, pharmaceuticals, therapies, and interventions which may be effective in treating a disease or other phenotypic characteristic of the animal. For example, animal models may be exposed to a compound or agent suspected of exhibiting an ability to ameliorate disease symptoms, at a sufficient concentration and for a time sufficient to elicit such an amelioration of disease symptoms in the exposed animals. The response of the animals to the exposure may be monitored by assessing the reversal of disorders associated with the disease. Exposure may involve treating mother animals during gestation of the model animals described herein, thereby exposing embryos or fetuses to the compound or agent which may prevent or ameliorate the disease or phenotype. Neonatal, juvenile, and adult animals can also be exposed.

More particularly, using the animal models of the invention, specifically, transgenic mice, methods of identifying agents, including compounds are provided, preferably, on the basis of the ability to affect at least one phenotype associated with a disruption in a target gene. In one embodiment, the present invention provides a method of identifying agents having an effect on target gene or alternatively, target protein expression or function. The method includes measuring a physiological response of the animal, for example, to the agent, and comparing the physiological response of such animal to a control animal, wherein the physiological response of the animal comprising a disruption

-19-

in a target gene as compared to the control animal indicates the specificity of the agent. A "physiological response" is any biological or physical parameter of an animal which can be measured. Molecular assays (e.g., gene transcription, protein production and degradation rates), physical parameters (e.g., exercise physiology tests, measurement of various parameters of respiration,

5   measurement of heart rate or blood pressure, measurement of bleeding time, aPTT.T, or TT), and cellular assays (e.g.,. immunohistochemical assays of cell surface markers, or the ability of cells to aggregate or proliferate) can be used to assess a physiological response. The transgenic animals and cells of the present invention may by utilized as models for diseases, disorders, or conditions associated with phenotypes relating to a disruption in a target gene.

10   The present invention provides a unique animal model for testing and developing new treatments relating to the behavioral phenotypes. Analysis of the behavioral phenotype allows for the development of an animal model useful for testing, for instance, the efficacy of proposed genetic and pharmacological therapies for human genetic diseases, such as neurological, neuropsychological, or psychotic illnesses.

15   A statistical analysis of the various behaviors measured can be carried out using any conventional statistical program routinely used by those skilled in the art (such as, for example, "Analysis of Variance" or ANOVA). A "p" value of about 0.05 or less is generally considered to be statistically significant, although slightly higher p values may still be indicative of statistically significant differences. To statistically analyze abnormal behavior, a comparison is made between the

20   behavior of a transgenic animal (or a group thereof) to the behavior of a wild-type mouse (or a group thereof), typically under certain prescribed conditions. "Abnormal behavior" as used herein refers to behavior exhibited by an animal having a disruption in the target gene, e.g. transgenic animal, which differs from an animal without a disruption in the target gene, *e.g.* wild-type mouse. Abnormal behavior consists of any number of standard behaviors that can be objectively measured (or observed)

25   and compared. In the case of comparison, it is preferred that the change be statistically significant to confirm that there is indeed a meaningful behavioral difference between the knockout animal and the wild-type control animal. Examples of behaviors which may be measured or observed include, but are not limited to, ataxia, rapid limb movement, eye movement, breathing, motor activity, cognition, emotional behaviors, social behaviors, hyperactivity, hypersensitivity, anxiety, impaired learning,

30   abnormal reward behavior, and abnormal social interaction, such as aggression.

     A series of tests may be used to measure the behavioral phenotype of the animal models of the present invention, including neurological and neuropsychological tests to identify abnormal behavior. These tests may be used to measure abnormal behavior relating to, for example, learning and memory, eating, pain, aggression, sexual reproduction, anxiety, depression, schizophrenia, and

35   drug abuse. (*See, e.g.,* Crawley and Paylor, *Hormones and Behavior* 31:197-211 (1997)).

-20-

The social interaction test involves exposing a mouse to other animals in a variety of settings. The social behaviors of the animals (e.g., touching, climbing, sniffing, and mating) are subsequently evaluated. Differences in behaviors can then be statistically analyzed and compared (*See, e.g.*, S. E. File, *et al.*, *Pharmacol. Bioch. Behav.* 22:941-944 (1985); R. R. Holson, *Phys. Behav.* 37:239-247

5   (1986)). Examplary behavioral tests include the following.

The mouse startle response test typically involves exposing the animal to a sensory (typically auditory) stimulus and measuring the startle response of the animal (*see, e.g.*, M. A. Geyer, *et al.*, *Brain Res. Bull.* 25:485-498 (1990); Paylor and Crawley, *Psychopharmacology* 132:169-180 (1997)). A pre-pulse inhibition test can also be used, in which the percent inhibition (from a normal startle

10   response) is measured by "cueing" the animal first with a brief low-intensity pre-pulse prior to the startle pulse.

The electric shock test generally involves exposure to an electrified surface and measurement of subsequent behaviors such as, for example, motor activity, learning, social behaviors. The behaviors are measured and statistically analyzed using standard statistical tests. (*See, e.g.*, G. J. Kant, *et al.*,

15   *Pharm. Bioch. Behav.* 20:793-797 (1984); N. J. Leidenheimer, *et al.*, *Pharmacol. Bioch. Behav.* 30:351-355 (1988)).

The tail-pinch or immobilization test involves applying pressure to the tail of the animal and/or restraining the animal's movements. Motor activity, social behavior, and cognitive behavior are examples of the areas that are measured. (*See, e.g.*, M. Bertolucci D'Angie, *et al.*, *Neurochem.*

20   55:1208-1214 (1990)).

The novelty test generally comprises exposure to a novel environment and/or novel objects. The animal's motor behavior in the novel environment and/or around the novel object are measured and statistically analyzed. (*See, e.g.*, D. K. Reinstein, *et al.*, *Pharm. Bioch. Behav.* 17:193-202 (1982); B. Poucet, *Behav. Neurosci.* 103:1009-10016 (1989); R. R. Holson, *et al.*, *Phys. Behav.* 37:231-238

25   (1986)). This test may be used to detect visual processing deficiencies or defects.

The learned helplessness test involves exposure to stresses, for example, noxious stimuli, which cannot be affected by the animal's behavior. The animal's behavior can be statistically analyzed using various standard statistical tests. (*See, e.g.*, A. Leshner, *et al.*, *Behav. Neural Biol.* 26:497-501 (1979)).

30   Alternatively, a tail suspension test may be used, in which the "immobile" time of the mouse is measured when suspended "upside-down" by its tail. This is a measure of whether the animal struggles, an indicator of depression. In humans, depression is believed to result from feelings of a lack of control over one's life or situation. It is believed that a depressive state can be elicited in animals by repeatedly subjecting them to aversive situations over which they have no control. A

35   condition of "learned helplessness" is eventually reached, in which the animal will stop trying to change its circumstances and simply accept its fate. Animals that stop struggling sooner are believed

-21-

to be more prone to depression. Studies have shown that the administration of certain antidepressant drugs prior to testing increases the amount of time that animals struggle before giving up.

The Morris water-maze test comprises learning spatial orientations in water and subsequently measuring the animal's behaviors, such as, for example, by counting the number of incorrect choices.

5   The behaviors measured are statistically analyzed using standard statistical tests. (*See, e.g.*, E. M. Spruijt, *et al.*, *Brain Res.* 527:192-197 (1990)).

Alternatively, a Y-shaped maze may be used (see, e.g., McFarland, D.J., *Pharmacology, Biochemistry and Behavior* 32:723-726 (1989); Dellu, F., et al., *Neurobiology of Learning and Memory* 73:31-48 (2000)). The Y-maze is generally believed to be a test of cognitive ability. The

10   dimensions of each arm of the Y-maze can be, for example, approximately 40 cm x 8 cm x 20 cm, although other dimensions may be used. Each arm can also have, for example, sixteen equally spaced photobeams to automatically detect movement within the arms. At least two different tests can be performed using such a Y-maze. In a continuous Y-maze paradigm, mice are allowed to explore all three arms of a Y-maze for, e.g., approximately 10 minutes. The animals are continuously tracked

15   using photobeam detection grids, and the data can be used to measure spontaneous alteration and positive bias behavior. Spontaneous alteration refers to the natural tendency of a "normal" animal to visit the least familiar arm of a maze. An alternation is scored when the animal makes two consecutive turns in the same direction, thus representing a sequence of visits to the least recently entered arm of the maze. Position bias determines egocentrically defined responses by measuring the

20   animal's tendency to favor turning in one direction over another. Therefore, the test can detect differences in an animal's ability to navigate on the basis of allocentric or egocentric mechanisms. The two-trial Y-maze memory test measures response to novelty and spatial memory based on a free-choice exploration paradigm. During the first trial (acquisition), the animals are allowed to freely visit two arms of the Y-maze for, e.g., approximately 15 minutes. The third arm is blocked off during

25   this trial. The second trial (retrieval) is performed after an intertrial interval of, e.g., approximately 2 hours. During the retrieval trial, the blocked arm is opened and the animal is allowed access to all three arms for, e.g., approximately 5 minutes. Data are collected during the retrieval trial and analyzed for the number and duration of visits to each arm. Because the three arms of the maze are virtually identical, discrimination between novelty and familiarity is dependent on "environmental"

30   spatial cues around the room relative to the position of each arm. Changes in arm entry and duration of time spent in the novel arm in a transgenic animal model may be indicative of a role of that gene in mediating novelty and recognition processes.

The passive avoidance or shuttle box test generally involves exposure to two or more environments, one of which is noxious, providing a choice to be learned by the animal. Behavioral

35   measures include, for example, response latency, number of correct responses, and consistency of response. (*See, e.g.*, R. Ader, *et al.*, *Psychon. Sci.* 26:125-128 (1972); R. R. Holson, *Phys. Behav.*

-22-

37:221-230 (1986)). Alternatively, a zero-maze can be used. In a zero-maze, the animals can, for example, be placed in a closed quadrant of an elevated annular platform having, e.g., 2 open and 2 closed quadrants, and are allowed to explore for approximately 5 minutes. This paradigm exploits an approach-avoidance conflict between normal exploratory activity and an aversion to open spaces in

5   rodents. This test measures anxiety levels and can be used to evaluate the effectiveness of anti-anxiolytic drugs. The time spent in open quadrants versus closed quadrants may be recorded automatically, with, for example, the placement of photobeams at each transition site.

       The food avoidance test involves exposure to novel food and objectively measuring, for example, food intake and intake latency. The behaviors measured are statistically analyzed using

10   standard statistical tests. (*See, e.g.*, B. A. Campbell, et al., *J. Comp. Physiol. Psychol.* 67:15-22 (1969)).

       The elevated plus-maze test comprises exposure to a maze, without sides, on a platform, the animal's behavior is objectively measured by counting the number of maze entries and maze learning. The behavior is statistically analyzed using standard statistical tests. (*See, e.g.*, H. A. Baldwin, *et al.*,

15   *Brain Res. Bull*, 20:603-606 (1988)).

       The stimulant-induced hyperactivity test involves injection of stimulant drugs (*e.g.*, amphetamines, cocaine, PCP, and the like), and objectively measuring, for example, motor activity, social interactions, cognitive behavior. The animal's behaviors are statistically analyzed using standard statistical tests. (*See, e.g.*, P. B. S. Clarke, *et al.*, *Psychopharmacology* 96:511-520 (1988); P.

20   Kuczenski, *et al.*, *J. Neuroscience* 11:2703-2712 (1991)).

       The self-stimulation test generally comprises providing the mouse with the opportunity to regulate electrical and/or chemical stimuli to its own brain. Behavior is measured by frequency and pattern of self-stimulation. Such behaviors are statistically analyzed using standard statistical tests. (*See, e.g.*, S. Nassif, *et al.*, *Brain Res.*, 332:247-257 (1985); W. L. Isaac, *et al.*, *Behav. Neurosci.*

25   103:345-355 (1989)).

       The reward test involves shaping a variety of behaviors, e.g., motor, cognitive, and social, measuring, for example, rapidity and reliability of behavioral change, and statistically analyzing the behaviors measured. (*See, e.g.*, L. E. Jarrard, *et al.*, *Exp. Brain Res.* 61:519-530 (1986)).

       The DRL (differential reinforcement to low rates of responding) performance test involves

30   exposure to intermittent reward paradigms and measuring the number of proper responses, e.g., lever pressing. Such behavior is statistically analyzed using standard statistical tests. (*See, e.g.*, J. D. Sinden, *et al.*, *Behav. Neurosci.* 100:320-329 (1986); V. Nalwa, *et al.*, *Behav Brain Res.* 17:73-76 (1985); and A. J. Nonneman, *et al.*, *J. Comp. Physiol. Psych.* 95:588-602 (1981)).

       The spatial learning test involves exposure to a complex novel environment, measuring the

35   rapidity and extent of spatial learning, and statistically analyzing the behaviors measured. (*See, e.g.*, N. Pitsikas, *et al.*, *Pharm. Bioch. Behav.* 38:931-934 (1991); B. Poucet, *et al.*, *Brain Res.* 37:269-280

-23-

(1990); D. Christie, *et al.*, *Brain Res.* 37:263-268 (1990); and F. Van Haaren, *et al.*, *Behav. Neurosci.* 102:481-488 (1988)). Alternatively, an open-field (of) test may be used, in which the greater distance traveled for a given amount of time is a measure of the activity level and anxiety of the animal. When the open field is a novel environment, it is believed that an approach-avoidance situation is created, in

5   which the animal is "torn" between the drive to explore and the drive to protect itself. Because the chamber is lighted and has no places to hide other than the corners, it is expected that a "normal" mouse will spend more time in the corners and around the periphery than it will in the center where there is no place to hide. "Normal" mice will, however, venture into the central regions as they explore more and more of the chamber. It can then be extrapolated that especially anxious mice will

10   spend most of their time in the corners, with relatively little or no exploration of the central region, whereas bold (i.e., less anxious) mice will travel a greater distance, showing little preference for the periphery versus the central region.

The visual, somatosensory and auditory neglect tests generally comprise exposure to a sensory stimulus, objectively measuring, for example, orientating responses, and statistically

15   analyzing the behaviors measured. (See, e.g., J. M. Vargo, et al., Exp. Neurol. 102:199-209 (1988)).

The consummatory behavior test generally comprises feeding and drinking, and objectively measuring quantity of consumption. The behavior measured is statistically analyzed using standard statistical tests. (See, e.g., P. J. Fletcher, et al., Psychopharmacol. 102:301-308 (1990); M. G. Corda, et al.,, Proc. Nat'l Acad. Sci. USA 80:2072-2076 (1983)).

20   A visual discrimination test can also be used to evaluate the visual processing of an animal. One or two similar objects are placed in an open field and the animal is allowed to explore for about 5-10 minutes. The time spent exploring each object (proximity to, i.e., movement within, e.g., about 3-5 cm of the object is considered exploration of an object) is recorded. The animal is then removed from the open field, and the objects are replaced by a similar object and a novel object. The animal is

25   returned to the open field and the percent time spent exploring the novel object over the old object is measured (again, over about a 5-10 minute span). "Normal" animals will typically spend a higher percentage of time exploring the novel object rather than the old object. If a delay is imposed between sampling and testing, the memory task becomes more hippocampal-dependent. If no delay is imposed, the task is more based on simple visual discrimination. This test can also be used for

30   olfactory discrimination, in which the objects (preferably, simple blocks) can be sprayed or otherwise treated to hold an odor. This test can also be used to determine if the animal can make gustatory discriminations; animals that return to the previously eaten food instead of novel food exhibit gustatory neophobia.

A hot plate analgesia test can be used to evaluate an animal's sensitivity to heat or painful

35   stimuli. For example, a mouse can be placed on an approximately 55°C hot plate and the mouse's response latency (e.g., time to pick up and lick a hind paw) can be recorded. These responses are not

-24-

reflexes, but rather "higher" responses requiring cortical involvement. This test may be used to evaluate a nociceptive disorder.

An accelerating rotarod test may be used to measure coordination and balance in mice. Animals can be, for example, placed on a rod that acts like a rotating treadmill (or rolling log). The rotarod can be made to rotate slowly at first and then progressively faster until it reaches a speed of, e.g., approximately 60 rpm. The mice must continually reposition themselves in order to avoid falling off. The animals are preferably tested in at least three trials, a minimum of 20 minutes apart. Those mice that are able to stay on the rod the longest are believed to have better coordination and balance.

A metrazol administration test can be used to screen animals for varying susceptibilities to seizures or similar events. For example, a 5mg/ml solution of metrazol can be infused through the tail vein of a mouse at a rate of, e.g., approximately 0.375 ml/min. The infusion will cause all mice to experience seizures, followed by death. Those mice that enter the seizure stage the soonest are believed to be more prone to seizures. Four distinct physiological stages can be recorded: soon after the start of infusion, the mice will exhibit a noticeable "twitch", followed by a series of seizures, ending in a final tensing of the body known as "tonic extension", which is followed by death.

Target Gene Products

The present invention further contemplates use of the target gene sequence to produce target gene products. Target gene products may include proteins that represent functionally equivalent gene products. Such an equivalent gene product may contain deletions, additions or substitutions of amino acid residues within the amino acid sequence encoded by the gene sequences described herein, but which result in a silent change, thus producing a functionally equivalent target gene product. Amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues involved.

For example, nonpolar (hydrophobic) amino acids include alanine, leucine, isoleucine, valine, proline, phenylalanine, tryptophan, and methionine; polar neutral amino acids include glycine, serine, threonine, cysteine, tyrosine, asparagine, and glutamine; positively charged (basic) amino acids include arginine, lysine, and histidine; and negatively charged (acidic) amino acids include aspartic acid and glutamic acid. "Functionally equivalent", as utilized herein, refers to a protein capable of exhibiting a substantially similar in vivo activity as the endogenous gene products encoded by the target gene sequences. Alternatively, when utilized as part of an assay, "functionally equivalent" may refer to peptides capable of interacting with other cellular or extracellular molecules in a manner substantially similar to the way in which the corresponding portion of the endogenous gene product would.

Other protein products useful according to the methods of the invention are peptides derived from or based on the target gene produced by recombinant or synthetic means (derived peptides).

-25-

Target gene products may be produced by recombinant DNA technology using techniques well known in the art. Thus, methods for preparing the gene polypeptides and peptides of the invention by expressing nucleic acid encoding gene sequences are described herein. Methods which are well known to those skilled in the art can be used to construct expression vectors containing gene

5   protein coding sequences and appropriate transcriptional/translational control signals. These methods include, for example, in vitro recombinant DNA techniques, synthetic techniques and in vivo recombination/genetic recombination (see, e.g., Sambrook, et al., 1989, supra, and Ausubel, et al., 1989, supra). Alternatively, RNA capable of encoding gene protein sequences may be chemically synthesized using, for example, automated synthesizers (see, e.g. Oligonucleotide Synthesis: A

10 Practical Approach, Gait, M. J. ed., IRL Press, Oxford (1984)).

A variety of host-expression vector systems may be utilized to express the gene coding sequences of the invention. Such host-expression systems represent vehicles by which the coding sequences of interest may be produced and subsequently purified, but also represent cells which may, when transformed or transfected with the appropriate nucleotide coding sequences, exhibit the gene

15 protein of the invention in situ. These include but are not limited to microorganisms such as bacteria (e.g., E. coli, B. subtilis) transformed with recombinant bacteriophage DNA, plasmid DNA or cosmid DNA expression vectors containing gene protein coding sequences; yeast (e.g. Saccharomyces, Pichia) transformed with recombinant yeast expression vectors containing the gene protein coding sequences; insect cell systems infected with recombinant virus expression vectors (e.g., baculovirus)

20 containing the gene protein coding sequences; plant cell systems infected with recombinant virus expression vectors (e.g., cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) or trans-formed with recombinant plasmid expression vectors (e.g., Ti plasmid) containing gene protein coding sequences; or mammalian cell systems (e.g. COS, CHO, BHK, 293, 3T3) harboring recombinant expression constructs containing promoters derived from the genome of mammalian

25 cells (e.g., metallothionine promoter) or from mammalian viruses (e.g., the adenovirus late promoter; the vaccinia virus 7.5 K promoter).

In bacterial systems, a number of expression vectors may be advantageously selected depending upon the use intended for the gene protein being expressed. For example, when a large quantity of such a protein is to be produced, for the generation of antibodies or to screen peptide

30 libraries, for example, vectors which direct the expression of high levels of fusion protein products that are readily purified may be desirable. Such vectors include, but are not limited, to the E. coli expression vector pUR278 (Ruther et al., EMBO J., 2:1791-94 (1983)), in which the gene protein coding sequence may be ligated individually into the vector in frame with the lac Z coding region so that a fusion protein is produced; pIN vectors (Inouye & Inouye, Nucleic Acids Res., 13:3101-09

35 (1985); Van Heeke et al., J. Biol. Chem., 264:5503-9 (1989)); and the like. pGEX vectors may also be used to express foreign polypeptides as fusion proteins with glutathione S-transferase (GST). In

-26-

general, such fusion proteins are soluble and can easily be purified from lysed cells by adsorption to glutathione-agarose beads followed by elution in the presence of free glutathione. The pGEX vectors are designed to include thrombin or factor Xa protease cleavage sites so that the cloned target gene protein can be released from the GST moiety.

5   In a preferred embodiment, full length cDNA sequences are appended with in-frame Bam HI sites at the amino terminus and Eco RI sites at the carboxyl terminus using standard PCR methodologies (Innis, et al. (eds) PCR Protocols: A Guide to Methods and Applications, Academic Press, San Diego (1990)) and ligated into the pGEX-2TK vector (Pharmacia, Uppsala, Sweden). The resulting cDNA construct contains a kinase recognition site at the amino terminus for radioactive labeling and
10   glutathione S-transferase sequences at the carboxyl terminus for affinity purification (Nilsson, *et al.*, *EMBO J.*, 4: 1075-80 (1985); Zabeau *et al.*, *EMBO J.*, 1: 1217-24 (1982)).

In an insect system, Autographa californica nuclear polyhedrosis virus (AcNPV) is used as a vector to express foreign genes. The virus grows in Spodoptera frugiperda cells. The gene coding sequence may be cloned individually into non-essential regions (for example the polyhedrin gene) of
15   the virus and placed under control of an AcNPV promoter (for example the polyhedrin promoter). Successful insertion of gene coding sequence will result in inactivation of the polyhedrin gene and production of non-occluded recombinant virus (*i.e.*, virus lacking the proteinaceous coat coded for by the polyhedrin gene). These recombinant viruses are then used to infect Spodoptera frugiperda cells in which the inserted gene is expressed (see, *e.g.*, Smith, *et al.*, *J. Virol.* 46: 584-93 (1983); U.S. Pat.
20   No. 4,745,051).

In mammalian host cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, the gene coding sequence of interest may be ligated to an adenovirus transcription/translation control complex, e.g., the late promoter and tripartite leader sequence. This chimeric gene may then be inserted in the adenovirus genome by in
25   vitro or in vivo recombination. Insertion in a non-essential region of the viral genome (*e.g.*, region E1 or E3) will result in a recombinant virus that is viable and capable of expressing gene protein in infected hosts. (*e.g.*, see Logan *et al.*, *Proc. Natl. Acad. Sci. USA*, 81:3655-59 (1984)). Specific initiation signals may also be required for efficient translation of inserted gene coding sequences. These signals include the ATG initiation codon and adjacent sequences. In cases where an entire
30   gene, including its own initiation codon and adjacent sequences, is inserted into the appropriate expression vector, no additional translational control signals may be needed. However, in cases where only a portion of the gene coding sequence is inserted, exogenous translational control signals, including, perhaps, the ATG initiation codon, must be provided. Furthermore, the initiation codon must be in phase with the reading frame of the desired coding sequence to ensure translation of the
35   entire insert. These exogenous translational control signals and initiation codons can be of a variety of origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of

appropriate transcription enhancer elements, transcription terminators, etc. (see Bitter, *et al.*, *Methods in Enzymol.*, 153:516-44 (1987)).

In addition, a host cell strain may be chosen which modulates the expression of the inserted sequences, or modifies and processes the gene product in the specific fashion desired. Such

5   modifications (*e.g.*, glycosylation) and processing (*e.g.*, cleavage) of protein products may be important for the function of the protein. Different host cells have characteristic and specific mechanisms for the post-translational processing and modification of proteins. Appropriate cell lines or host systems can be chosen to ensure the correct modification and processing of the foreign protein expressed. To this end, eukaryotic host cells which possess the cellular machinery for proper

10 processing of the primary transcript, glycosylation, and phosphorylation of the gene product may be used. Such mammalian host cells include but are not limited to CHO, VERO, BHK, HeLa, COS, MDCK, 293, 3T3, WI38, etc.

For long-term, high-yield production of recombinant proteins, stable expression is preferred. For example, cell lines which stably express the gene protein may be engineered. Rather than using

15 expression vectors which contain viral origins of replication, host cells can be transformed with DNA controlled by appropriate expression control elements (*e.g.*, promoter, enhancer, sequences, transcription terminators, polyadenylation sites, etc.), and a selectable marker. Following the introduction of the foreign DNA, engineered cells may be allowed to grow for 1-2 days in an enriched media, and then are switched to a selective media. The selectable marker in the recombinant plasmid

20 confers resistance to the selection and allows cells which stably integrate the plasmid into their chromosomes and grow, to form foci which in turn can be cloned and expanded into cell lines. This method may advantageously be used to engineer cell lines which express the gene protein. Such engineered cell lines may be particularly useful in screening and evaluation of compounds that affect the endogenous activity of the gene protein.

25 In a preferred embodiment, control of timing and/or quantity of expression of the recombinant protein can be controlled using an inducible expression construct. Inducible constructs and systems for inducible expression of recombinant proteins will be well known to those skilled in the art. Examples of such inducible promoters or other gene regulatory elements include, but are not limited to, tetracycline, metallothionine, ecdysone, and other steroid-responsive promoters, rapamycin

30 responsive promoters, and the like (No, *et al.*, *Proc. Natl. Acad. Sci. USA*, 93:3346-51 (1996); Furth, *et al.*, *Proc. Natl. Acad. Sci. USA*, 91:9302-6 (1994)). Additional control elements that can be used include promoters requiring specific transcription factors such as viral, particularly HIV, promoters. In one in embodiment, a Tet inducible gene expression system is utilized. (Gossen *et al.*, *Proc. Natl. Acad. Sci. USA*, 89:5547-51 (1992); Gossen, *et al.*, *Science*, 268:1766-69 (1995)). Tet Expression

35 Systems are based on two regulatory elements derived from the tetracycline-resistance operon of the *E. coli* Tn10 transposon—the tetracycline repressor protein (TetR) and the tetracycline operator

-28-

sequence (tetO) to which TetR binds. Using such a system, expression of the recombinant protein is placed under the control of the tetO operator sequence and transfected or transformed into a host cell. In the presence of TetR, which is co-transfected into the host cell, expression of the recombinant protein is repressed due to binding of the TetR protein to the tetO regulatory element. High-level,

5   regulated gene expression can then be induced in response to varying concentrations of tetracycline (Tc) or Tc derivatives such as doxycycline (Dox), which compete with tetO elements for binding to TetR. Constructs and materials for tet inducible gene expression are available commercially from CLONTECH Laboratories, Inc., Palo Alto, CA.

       When used as a component in an assay system, the gene protein may be labeled, either

10   directly or indirectly, to facilitate detection of a complex formed between the gene protein and a test substance. Any of a variety of suitable labeling systems may be used including but not limited to radioisotopes such as 125I; enzyme labeling systems that generate a detectable calorimetric signal or light when exposed to substrate; and fluorescent labels. Where recombinant DNA technology is used to produce the gene protein for such assay systems, it may be advantageous to engineer fusion

15   proteins that can facilitate labeling, immobilization and/or detection.

       Indirect labeling involves the use of a protein, such as a labeled antibody, which specifically binds to the gene product. Such antibodies include but are not limited to polyclonal, monoclonal, chimeric, single chain, Fab fragments and fragments produced by a Fab expression library.

       Production of Antibodies

20   Described herein are methods for the production of antibodies capable of specifically recognizing one or more epitopes. Such antibodies may include, but are not limited to polyclonal antibodies, monoclonal antibodies (mAbs), humanized or chimeric antibodies, single chain antibodies, Fab fragments, F(ab')2 fragments, fragments produced by a Fab expression library, anti-idiotypic (anti-Id) antibodies, and epitope-binding fragments of any of the above. Such antibodies may be used,

25   for example, in the detection of a target gene in a biological sample, or, alternatively, as a method for the inhibition of abnormal target gene activity. Thus, such antibodies may be utilized as part of disease treatment methods, and/or may be used as part of diagnostic techniques whereby patients may be tested for abnormal levels of target gene proteins, or for the presence of abnormal forms of such proteins.

30   For the production of antibodies, various host animals may be immunized by injection with the target gene, its expression product or a portion thereof. Such host animals may include but are not limited to rabbits, mice, and rats, to name but a few. Various adjuvants may be used to increase the immunological response, depending on the host species, including but not limited to Freund's (complete and incomplete), mineral gels such as aluminum hydroxide, surface active substances such

35   as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemocyanin,

-29-

dinitrophenol, and potentially useful human adjuvants such as BCG (*bacille Calmette-Guerin*) and *Corynebacterium parvum*.

Polyclonal antibodies are heterogeneous populations of antibody molecules derived from the sera of animals immunized with an antigen, such as target gene product, or an antigenic functional

5   derivative thereof. For the production of polyclonal antibodies, host animals such as those described above, may be immunized by injection with gene product supplemented with adjuvants as also described above.

Monoclonal antibodies, which are homogeneous populations of antibodies to a particular antigen, may be obtained by any technique which provides for the production of antibody molecules

10 by continuous cell lines in culture. These include, but are not limited to the hybridoma technique of Köhler and Milstein, *Nature*, 256:495-7 (1975); and U.S. Pat. No. 4,376,110), the human B-cell hybridoma technique (Kosbor, *et al.*, *Immunology Today*, 4:72 (1983); Cote, *et al.*, *Proc. Natl. Acad. Sci. USA*, 80:2026-30 (1983)), and the EBV-hybridoma technique (Cole, *et al.*, in Monoclonal Antibodies And Cancer Therapy, Alan R. Liss, Inc., New York, pp. 77-96 (1985)). Such antibodies

15 may be of any immunoglobulin class including IgG, IgM, IgE, IgA, IgD and any subclass thereof. The hybridoma producing the mAb of this invention may be cultivated in vitro or in vivo. Production of high titers of mAbs in vivo makes this the presently preferred method of production.

In addition, techniques developed for the production of "chimeric antibodies" (Morrison, *et al.*, *Proc. Natl. Acad. Sci.*, 81:6851-6855 (1984); Takeda, *et al.*, *Nature*, 314:452-54 (1985)) by

20 splicing the genes from a mouse antibody molecule of appropriate antigen specificity together with genes from a human antibody molecule of appropriate biological activity can be used. A chimeric antibody is a molecule in which different portions are derived from different animal species, such as those having a variable region derived from a murine mAb and a human immunoglobulin constant region.

25   Alternatively, techniques described for the production of single chain antibodies (U.S. Pat. No. 4,946,778; Bird, *Science* 242:423-26 (1988); Huston, *et al.*, *Proc. Natl. Acad. Sci. USA*, 85:5879-83 (1988); and Ward, *et al.*, *Nature*, 334:544-46 (1989)) can be adapted to produce gene-single chain antibodies. Single chain antibodies are formed by linking the heavy and light chain fragments of the Fv region via an amino acid bridge, resulting in a single chain polypeptide.

30   Antibody fragments which recognize specific epitopes may be generated by known techniques. For example, such fragments include but are not limited to: the F(ab')2 fragments which can be produced by pepsin digestion of the antibody molecule and the Fab fragments which can be generated by reducing the disulfide bridges of the F(ab')2 fragments. Alternatively, Fab expression libraries may be constructed (Huse, *et al.*, *Science*, 246:1275-81 (1989)) to allow rapid and easy

35 identification of monoclonal Fab fragments with the desired specificity.

-30-

Screening Methods

The present invention may be employed in a process for screening for agents such as agonists, *i.e.* agents that bind to and activate target gene polypeptides, or antagonists, *i.e.* inhibit the activity or interaction of target gene polypeptides with its ligand. Thus, polypeptides of the invention may also

5   be used to assess the binding of small molecule substrates and ligands in, for example, cells, cell-free preparations, chemical libraries, and natural product mixtures as known in the art. Any methods routinely used to identify and screen for agents that can modulate receptors may be used in accordance with the present invention.

The present invention provides methods for identifying and screening for agents that

10   modulate target gene expression or function. More particularly, cells that contain and express target gene sequences may be used to screen for therapeutic agents. Such cells may include non-recombinant monocyte cell lines, such as U937 (ATCC# CRL-1593), THP-1 (ATCC# TIB-202), and P388D1 (ATCC# TIB-63); endothelial cells such as HUVEC's and bovine aortic endothelial cells (BAEC's); as well as generic mammalian cell lines such as HeLa cells and COS cells, *e.g.*, COS-7

15   (ATCC# CRL-1651). Further, such cells may include recombinant, transgenic cell lines. For example, the transgenic mice of the invention may be used to generate cell lines, containing one or more cell types involved in a disease, that can be used as cell culture models for that disorder. While cells, tissues, and primary cultures derived from the disease transgenic animals of the invention may be utilized, the generation of continuous cell lines is preferred. For examples of techniques which

20   may be used to derive a continuous cell line from the transgenic animals, *see* Small, *et al.*, *Mol. Cell Biol.*, 5:642-48 (1985).

Target gene sequences may be introduced into, and overexpressed in, the genome of the cell of interest. In order to overexpress a target gene sequence, the coding portion of the target gene sequence may be ligated to a regulatory sequence which is capable of driving gene expression in the

25   cell type of interest. Such regulatory regions will be well known to those of skill in the art, and may be utilized in the absence of undue experimentation. Target gene sequences may also be disrupted or underexpressed. Cells having target gene disruptions or underexpressed target gene sequences may be used, for example, to screen for agents capable of affecting alternative pathways which compensate for any loss of function attributable to the disruption or underexpression.

30   *In vitro* systems may be designed to identify compounds capable of binding the target gene products. Such compounds may include, but are not limited to, peptides made of D-and/or L-configuration amino acids (in, for example, the form of random peptide libraries; see e.g., Lam, *et al.*, *Nature*, 354:82-4 (1991)), phosphopeptides (in, for example, the form of random or partially degenerate, directed phosphopeptide libraries; see, e.g., Songyang, *et al.*, *Cell*, 72:767-78 (1993)), antibodies, and

35   small organic or inorganic molecules. Compounds identified may be useful, for example, in modulating the activity of target gene proteins, preferably mutant target gene proteins; elaborating the

-31-

biological function of the target gene protein; or screening for compounds that disrupt normal target gene interactions or themselves disrupt such interactions.

The principle of the assays used to identify compounds that bind to the target gene protein involves preparing a reaction mixture of the target gene protein and the test compound under

5   conditions and for a time sufficient to allow the two components to interact and bind, thus forming a complex which can be removed and/or detected in the reaction mixture. These assays can be conducted in a variety of ways. For example, one method to conduct such an assay would involve anchoring the target gene protein or the test substance onto a solid phase and detecting target protein/ test substance complexes anchored on the solid phase at the end of the reaction. In one embodiment

10   of such a method, the target gene protein may be anchored onto a solid surface, and the test compound, which is not anchored, may be labeled, either directly or indirectly.

In practice, microtitre plates are conveniently utilized. The anchored component may be immobilized by non-covalent or covalent attachments. Non-covalent attachment may be accom- plished simply by coating the solid surface with a solution of the protein and drying. Alternatively, an

15   immobilized antibody, preferably a monoclonal antibody, specific for the protein may be used to anchor the protein to the solid surface. The surfaces may be prepared in advance and stored.

In order to conduct the assay, the nonimmobilized component is added to the coated surface containing the anchored component. After the reaction is complete, unreacted components are removed (e.g., by washing) under conditions such that any complexes formed will remain

20   immobilized on the solid surface. The detection of complexes anchored on the solid surface can be accomplished in a number of ways. Where the previously nonimmobilized component is pre-labeled, the detection of label immobilized on the surface indicates that complexes were formed. Where the previously nonimmobilized component is not pre-labeled, an indirect label can be used to detect complexes anchored on the surface; e.g., using a labeled antibody specific for the previously nonim-

25   mobilized component (the antibody, in turn, may be directly labeled or indirectly labeled with a labeled anti-Ig antibody).

Alternatively, a reaction can be conducted in a liquid phase, the reaction products separated from unreacted components, and complexes detected; e.g., using an immobilized antibody specific for target gene product or the test compound to anchor any complexes formed in solution, and a labeled

30   antibody specific for the other component of the possible complex to detect anchored complexes.

Compounds that are shown to bind to a particular target gene product through one of the methods described above can be further tested for their ability to elicit a biochemical response from the target gene protein. Agonists, antagonists and/or inhibitors of the expression product can be identified utilizing assays well known in the art.

35   Antisense, Ribozymes, and Antibodies

-32-

Other agents which may be used as therapeutics include the target gene, its expression product(s) and functional fragments thereof. Additionally, agents which reduce or inhibit mutant target gene activity may be used to ameliorate disease symptoms. Such agents include antisense, ribozyme, and triple helix molecules. Techniques for the production and use of such molecules are well known to those of skill in the art.

Anti-sense RNA and DNA molecules act to directly block the translation of mRNA by hybridizing to targeted mRNA and preventing protein translation. With respect to antisense DNA, oligodeoxyribonucleotides derived from the translation initiation site, e.g., between the -10 and +10 regions of the target gene nucleotide sequence of interest, are preferred.

Ribozymes are enzymatic RNA molecules capable of catalyzing the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by an endonucleolytic cleavage. The composition of ribozyme molecules must include one or more sequences complementary to the target gene mRNA, and must include the well known catalytic sequence responsible for mRNA cleavage. For this sequence, see U.S. Pat. No. 5,093,246, which is incorporated by reference herein in its entirety. As such within the scope of the invention are engineered hammerhead motif ribozyme molecules that specifically and efficiently catalyze endonucleolytic cleavage of RNA sequences encoding target gene proteins.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the molecule of interest for ribozyme cleavage sites which include the following sequences, GUA, GUU and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides corresponding to the region of the target gene containing the cleavage site may be evaluated for predicted structural features, such as secondary structure, that may render the oligonucleotide sequence unsuitable. The suitability of candidate sequences may also be evaluated by testing their accessibility to hybridization with complementary oligonucleotides, using ribonuclease protection assays.

Nucleic acid molecules to be used in triple helix formation for the inhibition of transcription should be single stranded and composed of deoxyribonucleotides. The base composition of these oligonucleotides must be designed to promote triple helix formation via Hoogsteen base pairing rules, which generally require sizeable stretches of either purines or pyrimidines to be present on one strand of a duplex. Nucleotide sequences may be pyrimidine-based, which will result in TAT and CGC triplets across the three associated strands of the resulting triple helix. The pyrimidine-rich molecules provide base complementarity to a purine-rich region of a single strand of the duplex in a parallel orientation to that strand. In addition, nucleic acid molecules may be chosen that are purine-rich, for example, containing a stretch of G residues. These molecules will form a triple helix with a DNA

-33-

duplex that is rich in GC pairs, in which the majority of the purine residues are located on a single strand of the targeted duplex, resulting in GGC triplets across the three strands in the triplex.

Alternatively, the potential sequences that can be targeted for triple helix formation may be increased by creating a so called "switchback" nucleic acid molecule. Switchback molecules are synthesized in an alternating 5'-3', 3'-5' manner, such that they base pair with first one strand of a duplex and then the other, eliminating the necessity for a sizeable stretch of either purines or pyrimidines to be present on one strand of a duplex.

It is possible that the antisense, ribozyme, and/or triple helix molecules described herein may reduce or inhibit the transcription (triple helix) and/or translation (antisense, ribozyme) of mRNA produced by both normal and mutant target gene alleles. In order to ensure that substantially normal levels of target gene activity are maintained, nucleic acid molecules that encode and express target gene polypeptides exhibiting normal activity may be introduced into cells that do not contain sequences susceptible to whatever antisense, ribozyme, or triple helix treatments are being utilized. Alternatively, it may be preferable to coadminister normal target gene protein into the cell or tissue in order to maintain the requisite level of cellular or tissue target gene activity.

Anti-sense RNA and DNA, ribozyme, and triple helix molecules of the invention may be prepared by any method known in the art for the synthesis of DNA and RNA molecules. These include techniques for chemically synthesizing oligodeoxyribonucleotides and oligoribonucleotides well known in the art such as for example solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding the antisense RNA molecule. Such DNA sequences may be incorporated into a wide variety of vectors which incorporate suitable RNA polymerase promoters such as the T7 or SP6 polymerase promoters. Alternatively, antisense cDNA constructs that synthesize antisense RNA constitutively or inducibly, depending on the promoter used, can be introduced stably into cell lines.

Various well-known modifications to the DNA molecules may be introduced as a means of increasing intracellular stability and half-life. Possible modifications include but are not limited to the addition of flanking sequences of ribonucleotides or deoxyribonucleotides to the 5' and/or 3' ends of the molecule or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the oligodeoxyribonucleotide backbone.

Antibodies that are both specific for target gene protein, and in particular, mutant gene protein, and interfere with its activity may be used to inhibit mutant target gene function. Such antibodies may be generated against the proteins themselves or against peptides corresponding to portions of the proteins using standard techniques known in the art and as also described herein. Such antibodies include but are not limited to polyclonal, monoclonal, Fab fragments, single chain antibodies, chimeric antibodies, etc.

-34-

In instances where the target gene protein is intracellular and whole antibodies are used, internalizing antibodies may be preferred. However, lipofectin liposomes may be used to deliver the antibody or a fragment of the Fab region which binds to the target gene epitope into cells. Where fragments of the antibody are used, the smallest inhibitory fragment which binds to the target or

5   expanded target protein's binding domain is preferred. For example, peptides having an amino acid sequence corresponding to the domain of the variable region of the antibody that binds to the target gene protein may be used. Such peptides may be synthesized chemically or produced via recombinant DNA technology using methods well known in the art (see, *e.g.*, Creighton, Proteins: Structures and Molecular Principles (1984) W.H. Freeman, New York 1983, *supra*; and Sambrook, *et*

10   *al.*, 1989, supra). Alternatively, single chain neutralizing antibodies which bind to intracellular target gene epitopes may also be administered. Such single chain antibodies may be administered, for example, by expressing nucleotide sequences encoding single-chain antibodies within the target cell population by utilizing, for example, techniques such as those described in Marasco, *et al.*, *Proc. Natl. Acad. Sci. USA*, 90:7889-93 (1993).

15   RNA sequences encoding *TARGET* gene protein may be directly administered to a patient exhibiting disease symptoms, at a concentration sufficient to produce a level of target gene protein such that disease symptoms are ameliorated. Patients may be treated by gene replacement therapy. One or more copies of a normal target gene, or a portion of the gene that directs the production of a normal target gene protein with target gene function, may be inserted into cells using vectors which

20   include, but are not limited to adenovirus, adeno-associated virus, and retrovirus vectors, in addition to other particles that introduce DNA into cells, such as liposomes. Additionally, techniques such as those described above may be utilized for the introduction of normal target gene sequences into human cells.

Cells, preferably, autologous cells, containing normal target gene expressing gene sequences

25   may then be introduced or reintroduced into the patient at positions which allow for the amelioration of disease symptoms.

Pharmaceutical Compositions, Effective Dosages, and Routes of Administration

The identified compounds that inhibit target mutant gene expression, synthesis and/or activity can be administered to a patient at therapeutically effective doses to treat or ameliorate the disease. A

30   therapeutically effective dose refers to that amount of the compound sufficient to result in amelioration of symptoms of the disease.

Toxicity and therapeutic efficacy of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, e.g., for determining the LD_{50} (the dose lethal to 50% of the population) and the ED_{50} (the dose therapeutically effective in 50% of the

35   population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio LD_{50}/ED_{50}. Compounds which exhibit large therapeutic indices are preferred.

-35-

While compounds that exhibit toxic side effects may be used, care should be taken to design a delivery system that targets such compounds to the site of affected tissue in order to minimize potential damage to uninfected cells and, thereby, reduce side effects.

5   The data obtained from the cell culture assays and animal studies can be used in formulating a range of dosage for use in humans. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED_{50} with little or no toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilized. For any compound used in the method of the invention, the therapeutically effective dose can be estimated initially from cell culture assays. A dose may be formulated in animal models to achieve a circulating

10   plasma concentration range that includes the IC_{50} (i.e., the concentration of the test compound which achieves a half-maximal inhibition of symptoms) as determined in cell culture. Such information can be used to more accurately determine useful doses in humans. Levels in plasma may be measured, for example, by high performance liquid chromatography.

Pharmaceutical compositions for use in accordance with the present invention may be

15   formulated in conventional manner using one or more physiologically acceptable carriers or excipients. Thus, the compounds and their physiologically acceptable salts and solvates may be formulated for administration by inhalation or insufflation (either through the mouth or the nose) or oral, buccal, parenteral, topical, subcutaneous, intraperitoneal, intravenous, intrapleural, intraoccular, intraarterial, or rectal administration. It is also contemplated that pharmaceutical compositions may

20   be administered with other products that potentiate the activity of the compound and optionally, may include other therapeutic ingredients.

For oral administration, the pharmaceutical compositions may take the form of, for example, tablets or capsules prepared by conventional means with pharmaceutically acceptable excipients such as binding agents (e.g., pregelatinized maize starch, polyvinylpyrrolidone or hydroxypropyl

25   methylcellulose); fillers (e.g., lactose, microcrystalline cellulose or calcium hydrogen phosphate); lubricants (e.g., magnesium stearate, talc or silica); disintegrants (e.g., potato starch or sodium starch glycolate); or wetting agents (e.g., sodium lauryl sulphate). The tablets may be coated by methods well known in the art. Liquid preparations for oral administration may take the form of, for example, solutions, syrups or suspensions, or they may be presented as a dry product for constitution with water

30   or other suitable vehicle before use. Such liquid preparations may be prepared by conventional means with pharmaceutically acceptable additives such as suspending agents (e.g., sorbitol syrup, cellulose derivatives or hydrogenated edible fats); emulsifying agents (e.g., lecithin or acacia); non-aqueous vehicles (e.g., almond oil, oily esters, ethyl alcohol or fractionated vegetable oils); and preservatives (e.g., methyl or propyl-p-hydroxybenzoates or sorbic acid). The preparations may also contain buffer

35   salts, flavoring, coloring and sweetening agents as appropriate.

-36-

Preparations for oral administration may be suitably formulated to give controlled release of the active compound.

For buccal administration the compositions may take the form of tablets or lozenges formulated in conventional manner.

5   For administration by inhalation, the compounds for use according to the present invention are conveniently delivered in the form of an aerosol spray presentation from pressurized packs or a nebulizer, with the use of a suitable propellant, *e.g.*, dichlorodifluoromethane, trichlorofluoromethane, dichlorotetrafluoroethane, carbon dioxide or other suitable gas. In the case of a pressurized aerosol the dosage unit may be determined by providing a valve to deliver a metered amount. Capsules and

10   cartridges of e.g. gelatin for use in an inhaler or insufflator may be formulated containing a powder mix of the compound and a suitable powder base such as lactose or starch.

The compounds may be formulated for parenteral administration by injection, *e.g.*, by bolus injection or continuous infusion. Formulations for injection may be presented in unit dosage form, *e.g.*, in ampoules or in multi-dose containers, with an added preservative. The compositions may take

15   such forms as suspensions, solutions or emulsions in oily or aqueous vehicles, and may contain formulatory agents such as suspending, stabilizing and/or dispersing agents. Alternatively, the active ingredient may be in powder form for constitution with a suitable vehicle, *e.g.*, sterile pyrogen-free water, before use.

The compounds may also be formulated in rectal compositions such as suppositories or

20   retention enemas, *e.g.*, containing conventional suppository bases such as cocoa butter or other glycerides. Oral ingestion is possibly the easiest method of taking any medication. Such a route of administration, is generally simple and straightforward and is frequently the least inconvenient or unpleasant route of administration from the patient's point of view. However, this involves passing the material through the stomach, which is a hostile environment for many materials, including

25   proteins and other biologically active compositions. As the acidic, hydrolytic and proteolytic environment of the stomach has evolved efficiently to digest proteinaceous materials into amino acids and oligopeptides for subsequent anabolism, it is hardly surprising that very little or any of a wide variety of biologically active proteinaceous material, if simply taken orally, would survive its passage through the stomach to be taken up by the body in the small intestine. The result, is that many

30   proteinaceous medicaments must be taken in through another method, such as parenterally, often by subcutaneous, intramuscular or intravenous injection.

Pharmaceutical compositions may also include various buffers (*e.g.*, Tris, acetate, phosphate), solubilizers (*e.g.*, Tween, Polysorbate), carriers such as human serum albumin, preservatives (thimerosol, benzyl alcohol) and anti-oxidants such as ascorbic acid in order to stabilize

35   pharmaceutical activity. The stabilizing agent may be a detergent, such as tween-20, tween-80, NP-40 or Triton X-100. EBP may also be incorporated into particulate preparations of polymeric

-37-

compounds for controlled delivery to a patient over an extended period of time. A more extensive survey of components in pharmaceutical compositions is found in Remington's Pharmaceutical Sciences, 18th ed., A. R. Gennaro, ed., Mack Publishing, Easton, Pa. (1990).

5   In addition to the formulations described previously, the compounds may also be formulated as a depot preparation. Such long acting formulations may be administered by implantation (for example subcutaneously or intramuscularly) or by intramuscular injection. Thus, for example, the compounds may be formulated with suitable polymeric or hydrophobic materials (for example as an emulsion in an acceptable oil) or ion exchange resins, or as sparingly soluble derivatives, for example, as a sparingly soluble salt.

10   The compositions may, if desired, be presented in a pack or dispenser device which may contain one or more unit dosage forms containing the active ingredient. The pack may for example comprise metal or plastic foil, such as a blister pack. The pack or dispenser device may be accompanied by instructions for administration.

Diagnostics

15   A variety of methods may be employed to diagnose disease conditions associated with the target gene. Specifically, reagents may be used, for example, for the detection of the presence of target gene mutations, or the detection of either over or under expression of target gene mRNA.

According to the diagnostic and prognostic method of the present invention, alteration of the wild-type target gene locus is detected. In addition, the method can be performed by detecting the

20   wild-type target gene locus and confirming the lack of a predisposition or neoplasia. "Alteration of a wild-type gene" encompasses all forms of mutations including deletions, insertions and point mutations in the coding and noncoding regions. Deletions may be of the entire gene or only a portion of the gene. Point mutations may result in stop codons, frameshift mutations or amino acid substitutions. Somatic mutations are those which occur only in certain tissues, e.g., in the tumor

25   tissue, and are not inherited in the germline. Germline mutations can be found in any of a body's tissues and are inherited. If only a single allele is somatically mutated, an early neoplastic state is indicated. However, if both alleles are mutated, then a late neoplastic state may be indicated. The finding of gene mutations thus provides both diagnostic and prognostic information. A target gene allele which is not deleted (e.g., that found on the sister chromosome to a chromosome carrying a

30   target gene deletion) can be screened for other mutations, such as insertions, small deletions, and point mutations. Mutations found in tumor tissues may be linked to decreased expression of the target gene product. However, mutations leading to non-functional gene products may also be linked to a cancerous state. Point mutational events may occur in regulatory regions, such as in the promoter of the gene, leading to loss or diminution of expression of the mRNA. Point mutations may also abolish

35   proper RNA processing, leading to loss of expression of the target gene product, or a decrease in mRNA stability or translation efficiency.

-38-

One test available for detecting mutations in a candidate locus is to directly compare genomic target sequences from cancer patients with those from a control population. Alternatively, one could sequence messenger RNA after amplification, *e.g.*, by PCR, thereby eliminating the necessity of determining the exon structure of the candidate gene. Mutations from cancer patients falling outside

5   the coding region of the target gene can be detected by examining the non-coding regions, such as introns and regulatory sequences near or within the target gene. An early indication that mutations in noncoding regions are important may come from Northern blot experiments that reveal messenger RNA molecules of abnormal size or abundance in cancer patients as compared to control individuals.

The methods described herein may be performed, for example, by utilizing pre-packaged

10   diagnostic kits comprising at least one specific gene nucleic acid or anti-gene antibody reagent described herein, which may be conveniently used, *e.g.*, in clinical settings, to diagnose patients exhibiting disease symptoms or at risk for developing disease.

Any cell type or tissue, preferably monocytes, endothelial cells, or smooth muscle cells, in which the gene is expressed may be utilized in the diagnostics described below.

15   DNA or RNA from the cell type or tissue to be analyzed may easily be isolated using procedures which are well known to those in the art. Diagnostic procedures may also be performed *in situ* directly upon tissue sections (fixed and/or frozen) of patient tissue obtained from biopsies or resections, such that no nucleic acid purification is necessary. Nucleic acid reagents may be used as probes and/or primers for such *in situ* procedures (see, for example, Nuovo, PCR *In Situ*

20   Hybridization: Protocols and Applications, Raven Press, N.Y. (1992)).

Gene nucleotide sequences, either RNA or DNA, may, for example, be used in hybridization or amplification assays of biological samples to detect disease-related gene structures and expression. Such assays may include, but are not limited to, Southern or Northern analyses, restriction fragment length polymorphism assays, single stranded conformational polymorphism analyses, *in situ*

25   hybridization assays, and polymerase chain reaction analyses. Such analyses may reveal both quantitative aspects of the expression pattern of the gene, and qualitative aspects of the gene expression and/or gene composition. That is, such aspects may include, for example, point mutations, insertions, deletions, chromosomal rearrangements, and/or activation or inactivation of gene expression.

30   Preferred diagnostic methods for the detection of gene-specific nucleic acid molecules may involve for example, contacting and incubating nucleic acids, derived from the cell type or tissue being analyzed, with one or more labeled nucleic acid reagents under conditions favorable for the specific annealing of these reagents to their complementary sequences within the nucleic acid molecule of interest. Preferably, the lengths of these nucleic acid reagents are at least 9 to 30

35   nucleotides. After incubation, all non-annealed nucleic acids are removed from the nucleic acid:fingerprint molecule hybrid. The presence of nucleic acids from the fingerprint tissue which

-39-

have hybridized, if any such molecules exist, is then detected. Using such a detection scheme, the nucleic acid from the tissue or cell type of interest may be immobilized, for example, to a solid support such as a membrane, or a plastic surface such as that on a microtitre plate or polystyrene beads. In this case, after incubation, non-annealed, labeled nucleic acid reagents are easily removed.

5   Detection of the remaining, annealed, labeled nucleic acid reagents is accomplished using standard techniques well-known to those in the art.

Alternative diagnostic methods for the detection of gene-specific nucleic acid molecules may involve their amplification, e.g., by PCR (the experimental embodiment set forth in Mullis U.S. Pat. No. 4,683,202 (1987)), ligase chain reaction (Barany, *Proc. Natl. Acad. Sci. USA*, 88:189-93 (1991)),

10   self sustained sequence replication (Guatelli, *et al., Proc. Natl. Acad. Sci.* USA, 87:1874-78 (1990)), transcriptional amplification system (Kwoh, *et al., Proc. Natl. Acad. Sci. USA*, 86:1173-77 (1989)), Q-Beta Replicase (Lizardi *et al., Bio/Technology,* 6:1197 (1988)), or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques well known to those of skill in the art. These detection schemes are especially useful for the detection of

15   nucleic acid molecules if such molecules are present in very low numbers.

In one embodiment of such a detection scheme, a cDNA molecule is obtained from an RNA molecule of interest (*e.g.*, by reverse transcription of the RNA molecule into cDNA). Cell types or tissues from which such RNA may be isolated include any tissue in which wild type fingerprint gene is known to be expressed, including, but not limited, to monocytes, endothelium, and/or smooth

20   muscle. A sequence within the cDNA is then used as the template for a nucleic acid amplification reaction, such as a PCR amplification reaction, or the like. The nucleic acid reagents used as synthesis initiation reagents (*e.g.*, primers) in the reverse transcription and nucleic acid amplification steps of this method may be chosen from among the gene nucleic acid reagents described herein. The preferred lengths of such nucleic acid reagents are at least 15-30 nucleotides. For detection of the

25   amplified product, the nucleic acid amplification may be performed using radioactively or non-radioactively labeled nucleotides. Alternatively, enough amplified product may be made such that the product may be visualized by standard ethidium bromide staining or by utilizing any other suitable nucleic acid staining method.

Antibodies directed against wild type or mutant gene peptides may also be used as disease

30   diagnostics and prognostics. Such diagnostic methods, may be used to detect abnormalities in the level of gene protein expression, or abnormalities in the structure and/or tissue, cellular, or subcellular location of fingerprint gene protein. Structural differences may include, for example, differences in the size, electronegativity, or antigenicity of the mutant fingerprint gene protein relative to the normal fingerprint gene protein.

35   Protein from the tissue or cell type to be analyzed may easily be detected or isolated using techniques which are well known to those of skill in the art, including but not limited to western blot

-40-

analysis. For a detailed explanation of methods for carrying out western blot analysis. (*See, e.g.* Sambrook, *et al.* (1989) *supra*, at Chapter 18). The protein detection and isolation methods employed herein may also be such as those described in Harlow and Lane, for example, (Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (1988)).

5   Preferred diagnostic methods for the detection of wild type or mutant gene peptide molecules may involve, for example, immunoassays wherein fingerprint gene peptides are detected by their interaction with an anti-fingerprint gene-specific peptide antibody.

For example, antibodies, or fragments of antibodies useful in the present invention may be used to quantitatively or qualitatively detect the presence of wild type or mutant gene peptides. This

10   can be accomplished, for example, by immunofluorescence techniques employing a fluorescently labeled antibody (see below) coupled with light microscopic, flow cytometric, or fluorimetric detection. Such techniques are especially preferred if the fingerprint gene peptides are expressed on the cell surface.

The antibodies (or fragments thereof) useful in the present invention may, additionally, be

15   employed histologically, as in immunofluorescence or immunoelectron microscopy, for *in situ* detection of fingerprint gene peptides. *In situ* detection may be accomplished by removing a histological specimen from a patient, and applying thereto a labeled antibody of the present invention. The antibody (or fragment) is preferably applied by overlaying the labeled antibody (or fragment) onto a biological sample. Through the use of such a procedure, it is possible to determine not only the

20   presence of the fingerprint gene peptides, but also their distribution in the examined tissue. Using the present invention, those of ordinary skill will readily perceive that any of a wide variety of histological methods (such as staining procedures) can be modified in order to achieve such *in situ* detection.

Immunoassays for wild type, mutant, or expanded fingerprint gene peptides typically

25   comprise incubating a biological sample, such as a biological fluid, a tissue extract, freshly harvested cells, or cells which have been incubated in tissue culture, in the presence of a detectably labeled antibody capable of identifying fingerprint gene peptides, and detecting the bound antibody by any of a number of techniques well known in the art.

The biological sample may be brought in contact with and immobilized onto a solid phase

30   support or carrier such as nitrocellulose, or other solid support which is capable of immobilizing cells, cell particles or soluble proteins. The support may then be washed with suitable buffers followed by treatment with the detectably labeled gene-specific antibody. The solid phase support may then be washed with the buffer a second time to remove unbound antibody. The amount of bound label on solid support may then be detected by conventional means.

35   The terms "solid phase support or carrier" are intended to encompass any support capable of binding an antigen or an antibody. Well-known supports or carriers include glass, polystyrene,

-41-

polypropylene, polyethylene, dextran, nylon, amylases, natural and modified celluloses, polyacrylamides, gabbros, and magnetite. The nature of the carrier can be either soluble to some extent or insoluble for the purposes of the present invention. The support material may have virtually any possible structural configuration so long as the coupled molecule is capable of binding to an

5   antigen or antibody. Thus, the support configuration may be spherical, as in a bead, or cylindrical, as in the inside surface of a test tube, or the external surface of a rod. Alternatively, the surface may be flat such as a sheet, test strip, etc. Preferred supports include polystyrene beads. Those skilled in the art will know many other suitable carriers for binding antibody or antigen, or will be able to ascertain the same by use of routine experimentation.

10   The binding activity of a given lot of anti-wild type or -mutant fingerprint gene peptide antibody may be determined according to well known methods. Those skilled in the art will be able to determine operative and optimal assay conditions for each determination by employing routine experimentation.

One of the ways in which the gene peptide-specific antibody can be detectably labeled is by

15   linking the same to an enzyme and using it in an enzyme immunoassay (EIA) (Voller, *Ric Clin Lab*, 8:289-98 (1978) ["The Enzyme Linked Immunosorbent Assay (ELISA)", Diagnostic Horizons 2:1-7, 1978, Microbiological Associates Quarterly Publication, Walkersville, Md.]; Voller, *et al., J. Clin. Pathol.*, 31:507-20 (1978); Butler, *Meth. Enzymol.*, 73:482-523 (1981); Maggio (ed.), Enzyme Immunoassay, CRC Press, Boca Raton, Fla. (1980); Ishikawa, *et al.,* (eds.) Enzyme Immunoassay,

20   Igaku-Shoin, Tokyo (1981)). The enzyme which is bound to the antibody will react with an appropriate substrate, preferably a chromogenic substrate, in such a manner as to produce a chemical moiety which can be detected, for example, by spectrophotometric, fluorimetric or by visual means. Enzymes which can be used to detectably label the antibody include, but are not limited to, malate dehydrogenase, staphylococcal nuclease, delta-5-steroid isomerase, yeast alcohol dehydrogenase,

25   alpha-glycerophosphate, dehydrogenase, triose phosphate isomerase, horseradish peroxidase, alkaline phosphatase, asparaginase, glucose oxidase, beta-galactosidase, ribonuclease, urease, catalase, glucose-6-phosphate dehydrogenase, glucoamylase and acetylcholinesterase. The detection can be accomplished by colorimetric methods which employ a chromogenic substrate for the enzyme. Detection may also be accomplished by visual comparison of the extent of enzymatic reaction of a

30   substrate in comparison with similarly prepared standards.

Detection may also be accomplished using any of a variety of other immunoassays. For example, by radioactively labeling the antibodies or antibody fragments, it is possible to detect fingerprint gene wild type, mutant, or expanded peptides through the use of a radioimmunoassay (RIA) (see, *e.g.,* Weintraub, B., Principles of Radioimmunoassays, Seventh Training Course on

35   Radioligand Assay Techniques, The Endocrine Society, March, 1986). The radioactive isotope can be

-42-

detected by such means as the use of a gamma counter or a scintillation counter or by autoradi-
ography.

It is also possible to label the antibody with a fluorescent compound. When the fluorescently
labeled antibody is exposed to light of the proper wave length, its presence can then be detected due to
5   fluorescence. Among the most commonly used fluorescent labeling compounds are fluorescein
isothiocyanate, rhodamine, phycoerythrin, phycocyanin, allophycocyanin, o-phthaldehyde and
fluorescamine.

The antibody can also be detectably labeled using fluorescence emitting metals such as
152Eu, or others of the lanthanide series. These metals can be attached to the antibody using such
10   metal chelating groups as diethylenetriaminepentacetic acid (DTPA) or ethylenediamine-tetraacetic
acid (EDTA).

The antibody also can be detectably labeled by coupling it to a chemiluminescent compound.
The presence of the chemiluminescent-tagged antibody is then determined by detecting the presence
of luminescence that arises during the course of a chemical reaction. Examples of particularly useful
15   chemiluminescent labeling compounds are luminol, isoluminol, theromatic acridinium ester,
imidazole, acridinium salt and oxalate ester.

Likewise, a bioluminescent compound may be used to label the antibody of the present
invention. Bioluminescence is a type of chemiluminescence found in biological systems in, which a
catalytic protein increases the efficiency of the chemiluminescent reaction. The presence of a
20   bioluminescent protein is determined by detecting the presence of luminescence. Important
bioluminescent compounds for purposes of labeling are luciferin, luciferase and aequorin.

Throughout this application, various publications, patents and published patent applications
are referred to by an identifying citation. The disclosures of these publications, patents and published
patent specifications referenced in this application are hereby incorporated by reference into the
25   present disclosure to more fully describe the state of the art to which this invention pertains.

The following examples are intended only to illustrate the present invention and should in no
way be construed as limiting the subject invention.

Examples

Example 1: Generation and Analysis of Mice Comprising GPCR Gene Disruptions

30   *Targeting Construct for Platelet Activating Factor Receptor Gene.* To investigate the role
of genes encoding GPCRs, particularly, platelet activating factor receptor genes, disruptions in genes
comprising the sequence set forth in SEQ ID NO:1 were produced by homologous recombination.
More particularly, as shown in Figure 2A-2B, a specific targeting construct having the ability to
disrupt or modify genes, specifically comprising SEQ ID NO:1 was created using as the targeting
35   arms (homologous sequences) in the construct, the sequences identified herein as SEQ ID NO:3 and
SEQ ID NO:4.

-43-

Transgenic Mice. The targeting construct was introduced into ES cells by electroporation and chimeric mice were generated. The ES cells were derived from 129/Sv-+P+Mgf-SLJ/J mouse substrain. F1 mice were generated by breeding with C57BL/6 females. The resultant F1N0 hetero-zygotes were intercrossed to produce F2N0 homozygotes, or were backcrossed to C57BL/6 mice to

5   generate F1N1 heterozygotes. F2N1 homozygous mutant mice were produced by intercrossing F1N1 heterozygous males and females.

Phenotypic Analysis. The transgenic mice were analyzed for phenotypic changes. The homozygous mice demonstrated at least one of the following behavior phenotypes:

Homozygous mice spent significantly more time in the central region on the open field test

10   than the wild-type mice as shown in Figure 4 and the data presented in Table 1 below. This indicates that the homozygous mice may have less anxiety in comparison to the wild-type mice.

TABLE 1 – OPEN FIELD TEST: TIME SPENT IN CENTRAL REGION

Genotype	F Generation	N Generation	Family	SubFamily	Gene Name	Time (sec)	Std. Err.	Count
+/+	2	0	GPCR	Orphan GPCR	platelet activating receptor homolog	15.69	4.81	3
+/+	2	1	GPCR	Orphan GPCR	platelet activating receptor homolog	34.36	6.20	9
-/-	2	0	GPCR	Orphan GPCR	platelet activating receptor homolog	18.35	6.06	6
-/-	2	1	GPCR	Orphan GPCR	platelet activating receptor homolog	49.07	4.13	11

15   Homozygous mice had significantly longer latencies to fan or lick their hind paw on the hot plate test as shown in Figure 3 and the data presented in Table 2 below. More particularly, the homozygous mice displayed an increased response latency to lick or fan their hind paw on the hot plate test, indicating that the homozygous mice may have a higher pain threshold in comparison to the wild-type mice.

20

-44-

TABLE 2 – HOT PLATE TEST: LATENCY TIME TO HINDPAW LICKING

Genotype	F Generation	N Generation	Family	SubFamily	Gene Name	Time (sec)	Std. Err.	Count
+/+	2	0	GPCR	Orphan GPCR	platelet activating receptor homolog	15.55	1.31	6
+/+	2	1	GPCR	Orphan GPCR	platelet activating receptor homolog	12.06	1.66	12
-/-	2	0	GPCR	Orphan GPCR	platelet activating receptor homolog	19.50	4.74	8
-/-	2	1	GPCR	Orphan GPCR	platelet activating receptor homolog	23.41	4.50	11

Example 2: Generation and Analysis of Mice Comprising GPCR Gene Disruptions

Targeting Construct for PAF Receptor Gene. To investigate the role of genes encoding

5   GPCRs, particularly, PAF receptor genes, disruptions in genes comprising the sequence set forth in SEQ ID NO:4 were produced by homologous recombination. More particularly, as shown in Figure 6A-6B, a specific targeting construct having the ability to disrupt or modify genes, specifically comprising SEQ ID NO:4 was created using as the targeting arms (homologous sequences) in the construct, the sequences identified herein as SEQ ID NO:5 and SEQ ID NO:6.

10   **Example 3: Generation and Analysis of Mice Comprising LPR5 Gene Disruptions**

Targeting Construct. To investigate the role of LPR5, disruptions in genes comprising the sequence set forth in SEQ ID NO:7 were produced by homologous recombination. More particularly, as shown in Figure 9A-9C, a specific targeting construct having the ability to disrupt or modify genes, specifically comprising SEQ ID NO:7 was created using as the targeting arms (homologous

15   sequences) in the construct, the sequences identified herein as SEQ ID NO:9 and SEQ ID NO:10.

Transgenic Mice. The targeting construct was introduced into ES cells derived from the 129/OlaHsd mouse substrain by electroporation to generate chimeric mice. F1 mice were generated by breeding with C57BL/6 females. The resultant F1N0 heterozygotes were intercrossed to produce F2N0 homozygotes, or were backcrossed to C57BL/6 mice to generate F1N1 heterozygotes. F2N1

20   homozygous mutant mice were produced by intercrossing F1N1 heterozygous males and females.

Phenotypic Analysis. The transgenic mice were analyzed for phenotypic changes. The homozygous mice demonstrated eye abnormalities, including retinal regeneration.

-45-

Specifically, histopathology studies demonstrated that the eyes of the homozygous mice suffered from retinal degeneration, including bilateral retinal regeneration. In each homozygous mutant, at least one of the following retinal changes were present: retinal folds; thinning and vacuolation of the pigment epithelium layer; degeneration of photoreceptors; thinning, disorganiza-
5   tion, and pyknosis of the outer nuclear layer; thinning and disorganization of the outer plexiform layer, including juxtaposition of the photoreceptor nuclei and the bipolar cell or inner nuclear layer; disorganization of the inner nuclear layer; thinning of the inner plexiform layer; loss of ganglion cell nuclei, especially large ganglion cells; and, gliosis of the nerve fiber layer. The changes were general-ly more prominent in the outer layers of the retina (photoreceptor layers) and least pronounced in the
10   inner layers (inner nuclear layer, inner plexiform layer, ganglion cell layer, and nerve fiber layer).

Behavior Analysis: The homozygous mice demonstrated at least one of the following behavior phenotypes:

The homozygous mice spent significantly less time in the central region in the open field test, indicating possible increased anxiety as compared to wild-type mice. This is shown in Figure 10.
15   The homozygous mice displayed a decrease in total distance traveled in the open field test. The homozygous mutants were significantly hypoactive, in that they moved about and explored the open field less than wild-type mice. This is shown in Figure 11.

Example 4: Generation and Analysis of Mice Comprising Cerberus Gene Disruptions

Targeting Construct. To investigate the role of cerberus genes, disruptions in genes
20   comprising the sequence set forth in SEQ ID NO:11 were produced by homologous recombination. More particularly, as shown in Figure 13A-13B, a specific targeting construct having the ability to disrupt or modify genes, specifically comprising SEQ ID NO:11 was created using as the targeting arms (homologous sequences) in the construct, the sequences identified herein as SEQ ID NO:13 and SEQ ID NO:14.
25   *Transgenic Mice.* The targeting construct was introduced into ES cells derived from the 129/SvJ x 129/Sv-CP mouse by electroporation to generate chimeric mice. F1 mice were generated by breeding with C57BL/6 females. The resultant F1N0 heterozygotes were intercrossed to produce F2N0 homozygotes, or were backcrossed to C57BL/6 mice to generate F1N1 heterozygotes. F2N1 homozygous mutant mice were produced by intercrossing F1N1 heterozygous males and females.
30   *Phenotypic Analysis.* The transgenic mice were analyzed for phenotypic changes. The homozygous mice at least one of the following behavioral phenotypes:

The homozygous mice displayed several differences during the open field test. Homozygous mice displayed a decrease in their average velocity during episodes of movement, and total distance traveled, indicating decreased activity (e.g., hypoactivity). In addition, homozygous mutants
35   displayed an increase in the number of fecal boli deposited during the ten-minute test, suggesting increased anxiety or nervousness.

-46-

The homozygous mice displayed a difference during the tail suspension test. Homozygous mice were more active during the six-minute test resulting in a decrease in immobility time, suggesting that the homozygote mice exhibit a decreased susceptibility to depression (anti-depressive behavior phenotype).

5   In summary, homozygous mutant mice displayed at least one of the following behaviors:

1) A decrease in average velocity of movement during open field testing compared to wild-type mice;

2) A decrease in total distance traveled during open field testing compared to wild-type mice (see Table 3 below and Figure 14);

10   3) An increase in the number of fecal boli during open field testing compared to wild-type mice; and

4) A decrease in total time immobile during the tail suspension test (see Table 4 below and Figure 15).

TABLE 3 – OPEN FIELD TEST: TOTAL DISTANCE TRAVELED

Genotype	F Generation	N Generation	Family	SubFamily	Gene Name	Distance (cm)	Std. Err.	Count
+/+	2	0	Growth Factor Inhibitor	Cerberus	cerberus 1	543.60	131.34	9
+/+	2	1	Growth Factor Inhibitor	Cerberus	cerberus 1	934.27	98.25	10
-/-	2	0	Growth Factor Inhibitor	Cerberus	cerberus 1	476.52	135.97	9
-/-	2	1	Growth Factor Inhibitor	Cerberus	cerberus 1	596.52	120.70	10

15

TABLE 4 – TAIL SUSPENSION TEST: TOTAL TIME IMMOBILE

Gen type	F Generation	N Generation	Family	SubFamily	Gene Name	Time (sec)	Std. Err.	Count
+/+	2	0	Growth Factor Inhibitor	Cerberus	cerberus 1	130.89	21.83	9
+/+	2	1	Growth Factor Inhibitor	Cerberus	cerberus 1	143.48	23.00	10
-/-	2	0	Growth Factor Inhibitor	Cerberus	cerberus 1	119.76	14.99	9
-/-	2	1	Growth Factor Inhibitor	Cerberus	cerberus 1	81.13	16.33	10

Example 5: Generation and Analysis of Mice Comprising Brainiac Gene Disruptions

Targeting Construct. To investigate the role of brainiac genes, disruptions in genes
comprising the sequence set forth in SEQ ID NO:17 were produced by homologous recombination.
More particularly, a specific targeting construct having the ability to disrupt or modify genes,
specifically comprising SEQ ID NO:17 was created using as the targeting arms (homologous
sequences) in the construct, the sequences identified herein as SEQ ID NO:15 and SEQ ID NO:16.

Transgenic Mice. The targeting construct was introduced into ES cells derived from
129/OlaHsd mouse substrain by electroporation to generate chimeric mice. F1 mice were generated
by breeding with C57BL/6 females. The resultant F1N0 heterozygotes were intercrossed to produce
F2N0 homozygotes, or were backcrossed to C57BL/6 mice to generate F1N1 heterozygotes. F2N1
homozygous mutant mice were produced by intercrossing F1N1 heterozygous males and females.

As is apparent to one of skill in the art, various modifications of the above embodiments can
be made without departing from the spirit and scope of this invention. These modifications and
variations are within the scope of this invention.

-48-

We claim:

1. A targeting construct comprising:

      (a) a first polynucleotide sequence homologous to a target gene;

      (b) a second polynucleotide sequence homologous to the target gene; and

      (c) a selectable marker.

2. The targeting construct of claim 1, wherein the targeting construct further comprises a screening marker.

3. A method of producing a targeting construct for a target gene, the method comprising:

      (a) obtaining a first polynucleotide sequence homologous to a target gene;

      (b) obtaining a second polynucleotide sequence homologous to the target gene;

      (c) providing a vector comprising a selectable marker; and

      (d) inserting the first and second sequences into the vector, to produce the targeting construct.

4. A method of producing a targeting construct for a target gene, the method comprising:

      (a) providing a polynucleotide comprising a first sequence homologous to a first region of a target gene and a second sequence homologous to a second region of a target gene; and

      (b) inserting a positive selection marker in between the first and second sequences to form the targeting construct.

5. A cell comprising a disruption in a target gene.

6. The cell of claim 5, wherein the cell is a murine cell.

7. The cell of claim 6, wherein the murine cell is an embryonic stem cell.

8. A non-human transgenic animal comprising a disruption in a target gene.

9. A cell derived from the non-human transgenic animal of claim 8.

10. A method of producing a transgenic mouse comprising a disruption in a target gene, the method comprising:

      (a) introducing the targeting construct of claim 1 into a cell;

      (b) introducing the cell into a blastocyst;

      (c) implanting the resulting blastocyst into a pseudopregnant mouse, wherein said pseudopregnant mouse gives birth to a chimeric mouse; and

      (d) breeding the chimeric mouse to produce the transgenic mouse.

11. A method of identifying an agent that modulates the expression of a target gene, the method comprising:

      (a) providing a non-human transgenic animal comprising a disruption in the target gene;

      (b) administering an agent to the non-human transgenic animal; and

(c) determining whether the expression of the disrupted target gene in the non-human transgenic animal is modulated.

12. A method of identifying an agent that modulates the function of a target gene, the method comprising:

(a) providing a non-human transgenic animal comprising a disruption in the target gene;

(b) administering an agent to the non-human transgenic animal; and

(c) determining whether the function of the disrupted target gene in the non-human transgenic animal is modulated.

13. A method of identifying an agent that modulates the expression of a target gene, the method comprising:

(a) providing a cell comprising a disruption in a target gene;

(b) contacting the cell with an agent; and

(c) determining whether expression of the target gene is modulated.

14. A method of identifying an agent that modulates the function of a target gene, the method comprising:

(a) providing a cell comprising a disruption in a target gene;

(b) contacting the cell with an agent; and

(c) determining whether the function of the target gene is modulated.

15. The method of claim 13 or claim 14, wherein the cell is derived from the non-human transgenic animal of claim 8.

16. An agent identified by the method of claim 11, claim 12, claim 13, or claim 14.

17. A transgenic mouse comprising a disruption in a platelet activating receptor gene, wherein the transgenic mouse exhibits less anxious behavior as compared to wild-type mice.

18. A transgenic mouse comprising a disruption in a platelet activating receptor gene, wherein the transgenic mouse exhibits a higher pain threshold as compared to wild-type mice.

19. The transgenic mouse of claim 17, wherein the transgenic mouse is heterozygous for a disruption in a platelet activating receptor gene.

20. The transgenic mouse of claim 17, wherein the transgenic mouse is homozygous for a disruption in a platelet activating receptor gene.

21. A method of producing a transgenic mouse comprising a disruption in a platelet activating receptor gene, wherein the transgenic mouse exhibits at least one of the following behaviors: less anxiety as compared to wild-type mice or higher pain threshold as compared to wild-type mouse, the method comprising:

-50-

(a) introducing a disruption in a platelet activating receptor gene targeting construct into a cell;

(b) introducing the cell into a blastocyst;

(c) implanting the resulting blastocyst into a pseudopregnant mouse, wherein said pseudopregnant mouse gives birth to a chimeric mouse; and

(d) breeding the chimeric mouse to produce the transgenic mouse comprising a disruption in a disruption in a platelet activating receptor gene.

22. A cell derived from the transgenic mouse of claim 17, claim 18 or claim 21, wherein the cell comprises a disruption in a platelet activating receptor gene.

23. A method of identifying an agent which modulates platelet activating receptor expression, the method comprising:

(a) administering an agent to the transgenic mouse comprising a disruption in a platelet activating receptor gene; and

(b) determining whether the agent modulates platelet activating receptor expression in the transgenic mouse, wherein the agent modulates a phenotype associated with a disruption in a platelet activating receptor gene.

24. A method of identifying an agent which modulates a phenotype associated with a disruption in a platelet activating receptor gene, the method comprising:

(a) administering an agent to a transgenic mouse comprising a disruption in a platelet activating receptor gene; and

(b) determining whether the agent modulates the phenotype.

25. A method of identifying an agent which modulates platelet activating receptor expression, the method comprising:

(a) providing a cell comprising a disruption in a platelet activating receptor gene;

(b) contacting the cell with an agent; and

(c) determining whether the agent modulates platelet activating receptor gene expression, wherein the agent modulates a phenotype associated with a disruption in a platelet activating receptor gene.

26. A method of identifying an agent which modulates platelet activating receptor gene function, the method comprising:

(a) providing a cell comprising a disruption in a platelet activating receptor gene;

(b) contacting the cell with an agent; and

-51-

(c) determining whether the agent modulates platelet activating receptor gene function, wherein the agent modulates a phenotype associated with a disruption in a platelet activating receptor gene.

27. An agent identified by the method of claim 23, claim 24, claim 25, or claim 26.

28. A transgenic mouse comprising a disruption in a LPR5 gene, wherein the transgenic mouse exhibits an eye abnormality.

29. The transgenic mouse of claim 28, wherein the eye abnormality is a retinal abnormality.

30. The transgenic mouse of claim 29, wherein the retinal abnormality is characterized by retinal degeneration.

31. The transgenic mouse of claim 30 wherein the transgenic mouse is heterozygous for a disruption in an LPR5 gene.

32. The transgenic mouse of claim 31, wherein the transgenic mouse is homozygous for a disruption in an LPR5 gene.

33. A method of producing a transgenic mouse comprising a disruption in an LPR5 gene, wherein the transgenic mouse exhibits an eye abnormality, the method comprising:

(a) introducing an LPR5 targeting construct into a cell;

(b) introducing the cell into a blastocyst;

(c) implanting the resulting blastocyst into a pseudopregnant mouse, wherein said pseudopregnant mouse gives birth to a chimeric mouse; and

(d) breeding the chimeric mouse to produce the transgenic mouse comprising a disruption in an LPR5 gene.

34. A cell derived from the transgenic mouse of claim 28 or claim 33.

35. A method of identifying an agent that ameliorates an eye abnormality, the method comprising:

(a) administering an agent to the transgenic mouse comprising a disruption in a LPR5 gene; and

(b) determining whether the agent ameliorates the eye abnormality in the transgenic mouse.

36. The method of claim 35, wherein the eye abnormality is a retinal abnormality.

37. A method of identifying an agent which modulates LPR5 gene expression, the method comprising:

(a) administering an agent to the transgenic mouse comprising a disruption in a LPR5 gene; and

(b) determining whether the agent modulates LPR5 expression in the transgenic mouse, wherein the agent modulates a phenotype associated with a disruption in a LPR5 gene.

38. The method of claim 37, wherein the phenotype comprises an eye abnormality.

39. The method of claim 38, wherein the eye abnormality is a retinal abnormality.

40. A method of identifying an agent which modulates LPR5 gene expression, the method comprising:

-52-

(a) providing a cell comprising a disruption in a LPR5 gene;

(b) contacting the cell with the agent; and

(c) determining whether the agent modulates LPR5 gene expression, wherein the agent modulates a phenotype associated with a disruption in a LPR5 gene.

41. The method of claim 40, wherein the phenotype is an eye abnormality.

42. The method of claim 41, wherein the eye abnormality is a retinal abnormality.

43. A method of identifying an agent which modulates LPR5 gene function, the method comprising:

(a) providing a cell comprising disruption in a LPR5 gene;

(b) contacting the cell with an agent; and

(c) determining whether the agent modulates LPR5 gene function, wherein the agent modulates a phenotype associated with a disruption in a LPR5 gene.

44. The method of claim 43, wherein the phenotype is an eye abnormality.

45. The method of claim 44, wherein the eye abnormality is a retinal abnormality.

46. A method of identifying an agent which modulates a phenotype associated with a disruption in a LPR5 gene, the method comprising:

(a) administering an agent to a transgenic mouse comprising a disruption in a LPR5 gene; and

(b) determining whether the agent modulates the phenotype.

47. The method of claim 46, wherein the phenotype is an eye abnormality.

48. The method of claim 44, wherein the eye abnormality is a retinal abnormality.

49. An agent identified by the method of claim 37, claim 40, claim 43, or claim 46.

50. A transgenic mouse comprising a disruption in a cerberus gene, wherein the transgenic mouse exhibits at least one of the following behaviors: hypoactivity, anxiety, nervousness or a decreased susceptibility to depression.

51. The transgenic mouse of claim 50, wherein the transgenic mouse is heterozygous for a disruption in cerberus gene.

52. The transgenic mouse of claim 50, wherein the transgenic mouse is homozygous for a disruption in a cerberus gene.

53. A method of producing a transgenic mouse comprising a disruption in a cerberus gene, wherein the transgenic mouse exhibits at least one of the following behaviors: hypoactivity, anxiety, nervousness or a decreased susceptibility to depression, the method comprising:

(a) introducing a cerberus gene targeting construct into a cell;

(b) introducing the cell into a blastocyst;

-53-

(c) implanting the resulting blastocyst into a pseudopregnant mouse, wherein said pseudopregnant mouse gives birth to a chimeric mouse; and

(d) breeding the chimeric mouse to produce the transgenic mouse comprising a disruption in a cerberus gene.

54. A cell derived from the transgenic mouse of claim 50 or claim 53.

55. A method of identifying an agent that ameliorates hypoactive, anxious or nervous behavior, the method comprising:

(a) administering an agent to a transgenic mouse comprising a disruption in a cerberus gene; and

(b) determining whether the agent ameliorates behavior of the transgenic mouse.

56. A method of identifying an agent which modulates cerberus gene expression, the method comprising:

(a) administering an agent to the transgenic mouse comprising a disruption in a cerberus gene; and

(b) determining whether the agent modulates cerberus expression in the transgenic mouse, wherein the agent has an effect on hypoactive, anxious, or nervous behavior of the transgenic mouse.

57. A method of identifying an agent which modulates hypoactive, anxious or nervous behavior associated with a disruption in a cerberus gene, the method comprising:

(a) administering an agent to a transgenic mouse comprising a disruption in a cerberus gene; and

(b) determining whether the agent modulates behavior of the transgenic mouse.

58. An agent identified by the method of claim 55, claim 56 or claim 57.

ATTCGGCACTGAGGAGTTAATACACAGCTGCAAGATATACCGAATACAAGAAC
CTGGGTTTGCCAAAATGATATCAGCTGTCGTGTGGCTCATGGTCCTGCTGATC
ATGGTGCCGAACATGGTGATTCCCATCAAGGACATCAAAGAAAAGTCAAATGT
AGGTTGCATGGAGTTTAAAAAGGAGTTTGGAAGAAACTGGCATTTGCTAACAA
ACTTCATCTGTGTGGCAATATTTTTAAATTTCTCAGTCATCATTTTGATATCC
AACTTCCTTGCAATCAGACAACTCTATAGAAATAGAGACAACACAAACTATCC
AAGTGTGAAATCAGCTCTGCTCCACATTCTCTTGGTGACAGCTAGTTACATCA
TATGCTTTGTTCCCTACCACGCTGTCAGGATCCCATACACCCTCAGCCAGACA
GAGGTCATATCTGATTGCTCAACTAGGATCGCACTCTTCAAAGCCAAAGAGGC
CACACTGCTGCTGGCGGTGTCTAATTTGTGTTTTGATCCAATCCTGTACTATC
ATCTGTCAAAAGCCTTTCGACTAAAGGTCACCGAGACTTTCGCCTCACCTAAA
AAGAGCAAGCCTCTGGAAGAAAGATTGAGAAGTGAGAATGATGTGTAAAATGC
AGGCTTTGCACGTTAGTCCTTCTTCCTGACTGGACCATAAAGCTGGTCACAGT
TTTC (SEQ ID NO:1)

$FIGURE\ 1$

Construct

Gene:	17	**GI Number(s):**	1738456
Gene Family:	GPCR		1749042
Gene Subfamily:	Orphan GPCR		
Gene Sequence:	EST contig, Mouse		

underlined = deleted in targeting construct

[ ] = sequence flanking Neo insert in targeting construct

ATTCGGCACTGAGGAG [TTAATACACAGCTGCAAGATATACCGAATACAAGAACCTGGGT
TTGCCAAAATGATATCAGCTGTCGTGTGGCTCATGGTCCTGCTGATCATGGTGCCGAACA
TGGTGA] TTCCCATCAAGGACATCAAAGAAAAGTCAAATGTAGGTTGCATGGAGTTTAAA
AAGGAGTTTGGAAGAAACTGGCATTTGCTAACAAACTTCATCTGTGTGTGGCAATATTTTTA
AATTTCTCAGTCATCATTTTGATATCCAACTTCCTTGCAATCAGACAACTCTATAGAAAT
AGAGACAACACAAACTATCCAAGTGTGAAATCAGCTCTGCTCCACATTCTCTTGGTGACA
GCTAGTTACATCATATGCTTTGTTCCCTACCACGCTGTCAGGATCCCATACACCCTCAGC
CAGACAGAGGTCATATCTGATTGCTCAACTAGGATCGCACTCTTCAAAGCCAAAGAGGCC
ACACTGCTGCTGGCGGTGTCTAATTTGTGTTTTGATCCAATCCTGTACTATCATCTGTCA
AAAGCCTTTCGACTAAAGGTCACCGAGACTTTCGCCTCACCTAAAAAGAGCAAGCCTCTG
GAAGAAAGATTGAGAAGTGAGAATGATGTGTAAAATGCAGGCTTTGCACGTTAGTCCTTC
TTCCTG [ACTGGACCATAAAGCTGGTCACAGTTTTC]

Gene Sequence Structure *	126 bp	Sequence Deleted	664 bp

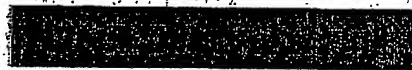Size of EST contig: 693 bp

FIG. 2-A

Targeting Vector* (genomic sequence)

Construct Number: 87

Arm Length:
5': 8 kb
3': 1.6 kb

——————— Targeting Vector
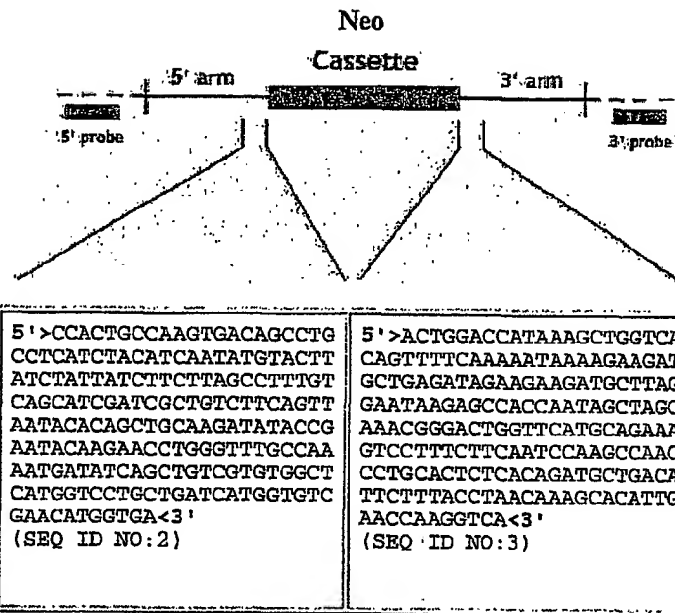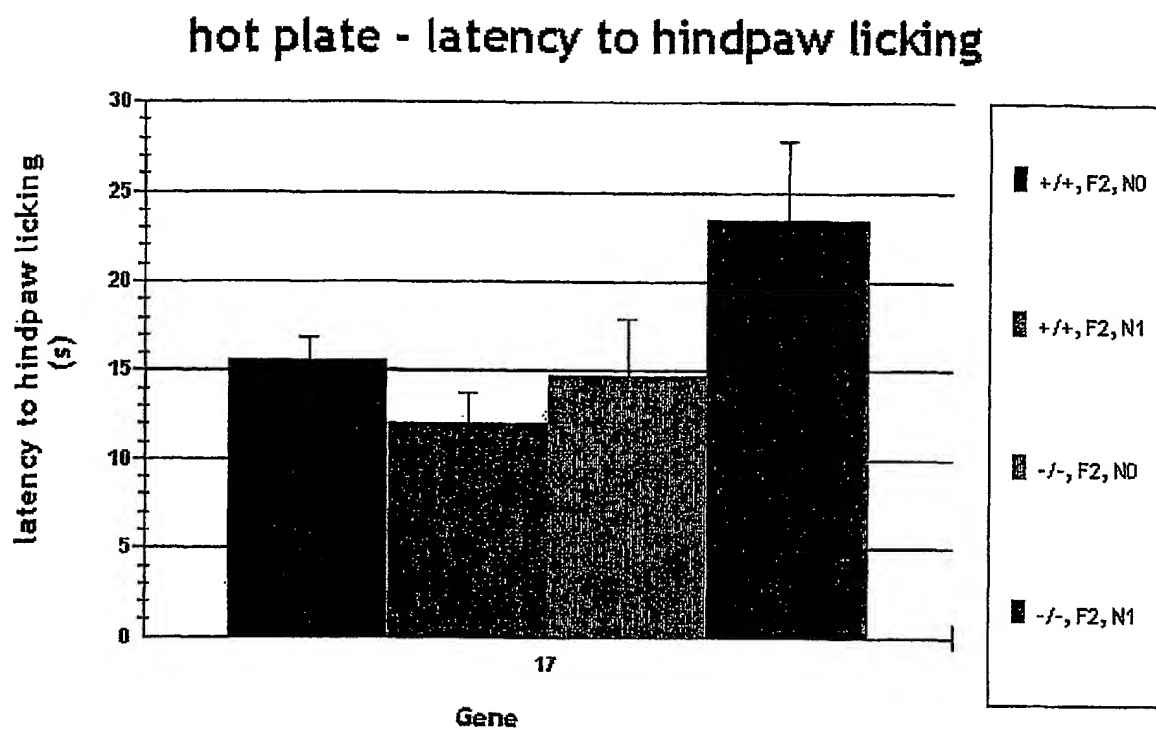— — — — Endogenous Locus

* Not drawn to scale


```
5'>CCACTGCCAAGTGACAGCCTG       5'>ACTGGACCATAAAGCTGGTCA
CCTCATCTACATCAATATGTACTT       CAGTTTTCAAAAATAAAAGAAGAT
ATCTATTATCTTCTTAGCCTTTGT       GCTGAGATAGAAGAAGATGCTTAG
CAGCATCGATCGCTGTCTTCAGTT       GAATAAGAGCCACCAATAGCTAGC
AATACACAGCTGCAAGATATACCG       AAACGGGACTGGTTCATGCAGAAA
AATACAAGAACCTGGGTTTGCCAA       GTCCTTTCTTCAATCCAAGCCAAC
AATGATATCAGCTGTCGTGTGGCT       CCTGCACTCTCACAGATGCTGACA
CATGGTCCTGCTGATCATGGTGTC       TTCTTTACCTAACAAAGCACATTG
GAACATGGTGA<3'               AACCAAGGTCA<3'
(SEQ ID NO:2)                  (SEQ ID NO:3)
```

FIG. 2B

hot plate - latency to hindpaw licking

Legend:
- +/+, F2, N0
- +/+, F2, N1
- -/-, F2, N0
- -/-, F2, N1

Y-axis: latency to hindpaw licking (s)

X-axis: Gene (17)

FIG. 3

Phenotypic Data Summary - Open Field

2,0,-/-,Male

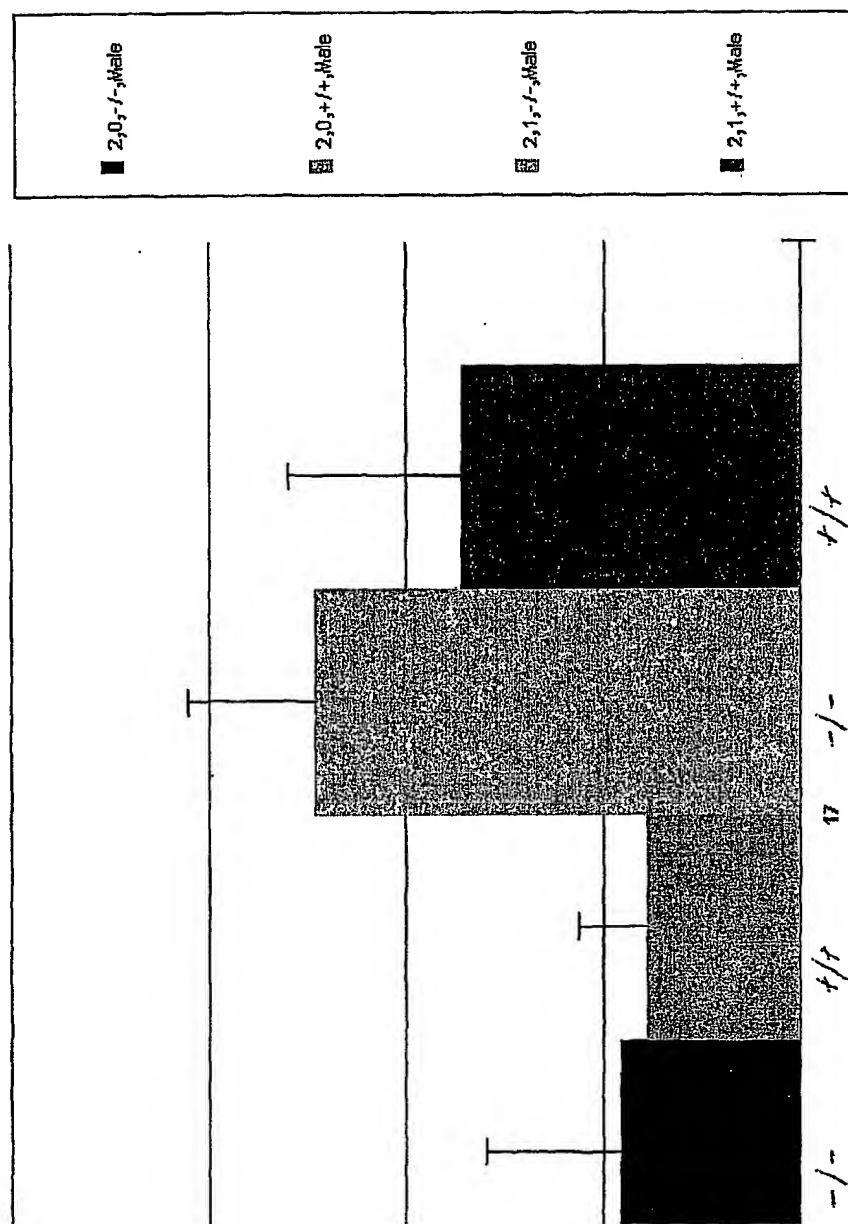2,0,+/+,Male

2,1,-/-,Male

2,1,+/+,Male

FIG. 4

GGCCTCTGAGAACCTTGGTGTGCCAAGTTACTTCAGTCACATGTTATTTTACA
ATGTATATCAGTATATCGTTCCTGGGGTTGATAACCATTGACCGCTACCTGAA
GACCACCAGGCCATTTAAAACGTCCAGCCCCAGCAATCTCTTGGGTGCAAAGA
TTCTTTCTGTTGTCATCTGGGCCTTCATGTTCTTAATTTCACTGCCTAACATG
ATTCTCACCAACAGGAGGCCAAAAGATAAGGACGTAACAAAATGTTCTTTCTT
AAAGTCAGAGTTTGGTCTAGTTTGGCACGAAATAGTCAATTACATCTGCCAAG
TCATTTTCTGGATTAATTTTTTAATTGTCATCGTTTGTTATAGCCTCATTACC
AAAGAACTCTATCGGTCTTATGTCAGAACAAGGGGTTCAGCCAAAGTTCCCAA
GAGAAAGGTAAACGTCAAGGTTTTCATCATCATTGCTGTATTCTTTATTTGCT
TTGTTCCCTTCCACTTTGCACGGATTCCCTACACCCTGAGGCAGACTCGGGCC
GTCTTTGACTGC (SEQ ID NO:4)

FIG. 5

Construct

Gene: 43

GI Number(s): 1912557

Gene Family: GPCR

Gene Subfamily: Orphan GPCR

Gene Sequence: EST, Mouse

<u>underlined</u> = deleted in targeting construct

[ ] = sequence flanking Neo insert in targeting construct

[GGCCTCTGAGAACCTTGGTGTGCCAAGTTACTTC]<u>AGTCACATGTTATTTTACAATGTA
TATCAGTATATCGTTCCTGGGGTTGATAACCATTGACCGCTACCTGAAGACCACCAGGCC
ATTTAAAACGTCCAGCCCCAGCAATCTCTTGGGTGCAAAGATTCTTTCTGTTGTCATCTG
GGCCTTCATGTTCTTAATTTCACTGCCTAACATGATTCTCACCAACAGGAGGCCAAAAGA
TAAGGACGTAACAAAATGTTCTTTCTTAAAGTCAGAGTTTGGTCTAGTTTGGCACGAAAT
AGTCAATTACATCTGCCAAGTCATTTTCTGGATTAATTTTTTAATTGTCATCGTTTGTTA
TAGCCTCATTACCAAAGAACTCTATCGGTCTTATGTC</u>[AGAACAAGGGGTTCAGCCAAAG
TTCCCAAGAGAAAGGTAAACGTCAAGGTTTTCATCATCATTGCTGTATTCTTTATTTGCT
TTGTTCCCTTCCACTTTGCACGGATTCCCTACACCCTGAGGCAGACTCGGGCCGTCTTTG
ACTGC]

Gene Sequence Structure *

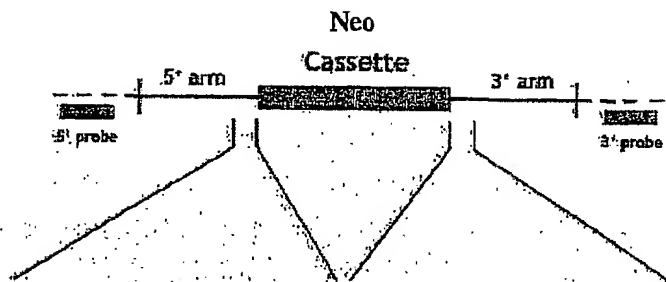| 35 bp | Sequence Deleted | 395 bp |

Size of EST: 542 bp

FIG. 6A

Targeting Vector* (genomic sequence)

Construct Number: 263

Arm Length:
5': 1.1 kb
3': 3.3 kb

————— Targeting Vector
— · — · — · — Endogenous Locus

* Not drawn to scale


```
5'>CTGTTCTTTGCTGGGCTCATC
ACGAACAGCTTGGCAATGAGGATT
TTCTTTCAGATCCGCAGTAAATCC
AACTTCATCATTTTTCTTAAGAAC
ACGGTCATCTCTGATCTACTAATG
ATTCTAACTTTTCCATTTAAAATT
CTTAGTGATGCTAAACTGGGAGCC
GGGCCTCTGAGAACCTTGGTGTGC
CAAGTTACTTC<3'
(SEQ ID NO:5)
```

```
5'>AGAACAAGGGGTTCAGCCAAA
GTTCCCAAGAAAAAGGTAAACGTC
AAGGTTTTCATCATCATTGCTGTA
TTCTTTATTTGCTTTGTTCCCTTC
CACTTTGCACGGATTCCCTACACC
CTGAGCCAAACTCGGGCCGTCTTT
GACTGCAGTGCTGAGAACACCCTG
TTCTACGTGAAGGAGAGCACCCTA
TGGCTGACGTC<3'
(SEQ ID NO:6)
```

FIG. 6.B

```
GCCGCGGCGCCCGAGGCGGGAGCAAGAGGCGCCGGGAGCCGCGAGGATCCACC
GCCGCCGCGCGCGCCATGGAGCCCGAGTGAGCGCGCGGCGCTCCCGGCCGCCG
GACGACATGGAAACGGCGCCGACCCGGGCCCCTCCGCCGCCGCCGCCGCCGCT
GCTGCTGCTGGTGCTGTACTGCAGCTTGGTCCCCGCCGCGGCCTCACCGCTCC
TGTTGTTTGCCAACCGCCGGGATGTGCGGCTAGTGGATGCCGGCGGAGTGAAG
CTGGAGTCCACCATTGTGGCCAGTGGCCTGGAGGATGCAGCTGCTGTAGACTT
CCAGTTCTCCAAGGGTGCTGTGTACTGGACAGATGTGAGCGAGGAGGCCATCA
AACAGACCTACCTGAACCAGACTGGAGCTGCTGCACAGAACATTGTCATCTCG
GGCCTCGTGTCACCTGATGGCCTGGCCTGTGACTGGGTTGGCAAGAAGCTGTA
CTGGACGGACTCCGAGACCAACCGCATTGAGGTTGCCAACCTCAATGGGACGT
CCCGTAAGGTTCTCTTCTGGCAGGACCTGGACCAGCCAAGGGCCATTGCCCTG
GATCCTGCACATGGGTACATGTACTGGACTGACTGGGGGGAAGCACCCCGGAT
CGAGCGGGCAGGGATGGATGGCAGTACCCGGAAGATCATTGTAGACTCCGACA
TTTACTGGCCCAATGGGCTGACCATCGACCTGGAGGAACAGAAGCTGTACTGG
GCCGATGCCAAGCTCAGCTTCATCCACCGTGCCAACCTGGACGGCTCCTTCCG
GCAGAAGGTGGTGGAGGGCAGCCTCACTCACCCTTTTGCCCTGACACTCTCTG
GGGACACACTCTACTGGACAGACTGGCAGACCCGCTCCATCCACGCCTGCAAC
AAGTGGACAGGGGAGCAGAGGAAGGAGATCCTTAGTGCTCTGTACTCACCCAT
GGACATCCAAGTGCTGAGCCAGGAGCGGCAGCCTCCCTTCCACACACCATGCG
AGGAGGACAACGGTGGCTGTTCCCACCTGTGCCTGCTGTCCCCGAGGGAGCCT
TTCTACTCCTGTGCCTGCCCCACTGGTGTGCAGTTGCAGGACAATGGCAAGAC
GTGCAAGACAGGGGCTGAGGAAGTGCTGCTGCTGGCTCGGAGGACAGACCTGA
GGAGGATCTCTCTGGACACCCCTGACTTCACAGACATAGTGCTGCAGGTGGGC
GACATCCGGCATGCCATTGCCATTGACTACGATCCCCTGGAGGGCTACGTGTA
CTGGACCGATGATGAGGTGCGGGCTATCCGCAGGGCGTACCTAGATGGCTCAG
GTGCGCAGACACTTGTGAACACTGAGATCAATGACCCCGATGGCATTGCTGTG
GACTGGGTCGCCCGGAACCTCTACTGGACAGATACAGGCACTGACAGAATTGA
GGTGACTCGCCTCAACGGCACCTCCCGAAAGATCCTGGTATCTGAGGACCTGG
ACGAACCGCGAGCCATTGTGTTGCACCCTGTGATGGGCCTCATGTACTGGACA
GACTGGGGGGAGAACCCCAAAATCGAATGCGCCAACCTAGATGGGAGAGATCG
GCATGTCCTGGTGAACACCTCCCTTGGGTGGCCCAATGGACTGGCCCTGGACC
TGCAGGAGGGCAAGCTGTACTGGGGGGATGCCAAAACTGATAAAATCGAGGTG
ATCAACATAGACGGGACAAAGCGGAAGACCCTGCTTGAGGACAAGCTCCCACA
CATTTTTGGGTTCACACTGCTGGGGGACTTCATCTACTGGACCGACTGGCAGA
GACGCAGTATTGAAAGGGTCCACAAGGTCAAGGCCAGCCGGGATGTCATCATT
GATCAACTCCCCGACCTGATGGGACTCAAAGCCGTGAATGTGGCCAAGGTTGT
CGGAACCAACCCATGTGCGGATGGAAATGGAGGGTGCAGCCATCTGTGCTTCT
TCACCCCACGTGCCACCAAGTGTGGCTGCCCCATTGGCCTGGAGCTGTTGAGT
GACATGAAGACCTGCATAATCCCCGAGGCCTTCCTGGTATTCACCAGCAGAGC
CACCATCCACAGGATCTCCCTGGAGACTAACAACAACGATGTGGCTATCCCAC
TCACGGGTGTCAAAGAGGCCTCTGCACTGGACTTTGATGTGTCCAACAATCAC
```

FIG. 7A

```
ATCTACTGGACTGATGTTAGCCTCAAGACGATCAGCCGAGCCTTCATGAATGG
GAGCTCAGTGGAGCACGTGATTGAGTTTGGCCTCGACTACCCTGAAGGAATGG
CTGTGGACTGGATGGGCAAGAACCTCTATTGGGCGGACACAGGGACCAACAGG
ATTGAGGTGGCCCGGCTGGATGGGCAGTTCCGGCAGGTGCTTGTGTGGAGAGA
CCTTGACAACCCCAGGTCTCTGGCTCTGGATCCTACTAAAGGCTACATCTACT
GGACTGAGTGGGGTGGCAAGCCAAGGATTGTGCGGGCCTTCATGGATGGGACC
AATTGTATGACACTGGTAGACAAGGTGGGCCGGGCCAACGACCTCACCATTGA
TTATGCCGACCAGCGACTGTACTGGACTGACCTGGACACCAACATGATTGAGT
CTTCCAACATGCTGGGTCAGGAGCGCATGGTGATAGCTGACGATCTGCCCTAC
CCGTTTGGCCTGACTCAATATAGCGATTACATCTACTGGACTGACTGGAACCT
GCATAGCATTGAACGGGCGGACAAGACCAGTGGGCGGAACCGCACCCTCATCC
AGGGTCACCTGGACTTCGTCATGGACATCCTGGTGTTCCACTCCTCCCGTCAG
GATGGCCTCAACGACTGCGTGCACAGCAATGGCCAGTGTGGGCAGCTGTGCCT
CGCCATCCCCGGAGGCCACCGCTGTGGCTGTGCTTCACACTACACGCTGGACC
CCAGCAGCCGCAACTGCAGCCCGCCCTCCACCTTCTTGCTGTTCAGCCAGAAA
TTTGCCATCAGCCGGATGATCCCCGATGACCAGCTCAGCCCGGACCTTGTCCT
ACCCCTTCATGGGCTGAGGAACGTCAAAGCCATCAACTATGACCCGCTGGACA
AGTTCATCTACTGGGTGGACGGGCGCCAGAACATCAAGAGGGCCAAGGACGAC
GGTACCCAGCCCTCCATGCTGACCTCTCCCAGCCAAAGCCTGAGCCCAGACAG
ACAGCCACGACCTCAGCATTGACATCTACAGCCGGACACTGTTCTGGACCT
GTGAGGCCACCAACACTATCAATGTCCACCGGCTGGATGGGGATGCCATGGGA
GTGGTGCTTCGAGGGGACCGTGACAAGCCAAGGGCCATTGCTGTCAATGCTGA
GCGAGGGTACATGTACTTTACCAACATGCAGGACCATGCTGCCAAGATCGAGC
GAGCCTCCCTGGATGGCACAGAGCGGGAGGTCCTCTTCACCACAGGCCTCATC
CGTCCCGTGGCCCTTGTGGTGGACAATGCTCTGGGCAAGCTCTTCTGGGTGGA
TGCCGACCTAAAGCGAATCGAAAGCTGTGACCTCTCTGGGGCCAACCGCCTGA
CCCTGGAAGATGCCAACATCGTACAGCCAGTAGGTCTGACAGTGCTGGGCAGG
CACCTCTACTGGATCGACCGCCAGCAGCAGATGATCGAGCGCGTGGAGAAGAC
CACTGGGGACAAGCGGACTAGGGTTCAGGGCCGTGTCACCCACCTGACAGGCA
TCCATGCCGTGGAGGAAGTCAGCCTGGAGGAGTTCTCAGCCCATCCTTGTGCC
CGAGACAATGGCGGCTGCTCCCACATCTGTATCGCCAAGGGTGATGGAACACC
GCGCTGCTCGTGCCCTGTCCACCTGGTGCTCCTGCAGAACCTGCTGACTTGTG
GTGAGCCTCCTACCTGCTCCCCTGATCAGTTTGCATGTACCACTGGTGAGATC
GACTGCATCCCCGGAGCCTGGCGCTGTGACGGCTTCCCTGAGTGTGCTGACCA
GAGTGATGAAGAAGGCTGCCCAGTGTGCTCCGCCTCTCAGTTCCCCTGCGCTC
GAGGCCAGTGTGTGGACCTGCGGTTACGCTGCGACGGTGAGGCCGACTGCCAG
GATCGCTCTGATGAAGCTAACTGCGATGCTGTCTGTCTGCCCAATCAGTTCCG
GTGCACCAGCGGCCAGTGTGTCCTCATCAAGCAACAGTGTGACTCCTTCCCCG
ACTGTGCTGATGGGTCTGATGAGCTCATGTGTGAAATCAACAAGCCACCCTCT
GATGACATCCCAGCCCACAGCAGTGCCATTGGGCCCGTCATTGGTATCATCCT
CTCCCTCTTCGTCATGGGCGGGGTCTACTTTGTCTGCCAGCGTGTGATGTGCC
```

AGCGCTACACAGGGGCCAGTGGGCCCTTTCCCCACGAGTATGTTGGTGGAGCC
CCTCATGTGCCTCTCAACTTCATAGCCCCAGGTGGCTCACAGCACGGTCCCTT
CCCAGGCATCCCGTGCAGCAAGTCCGTGATGAGCTCCATGAGCCTGGTGGGGG
GGCGCGGCAGCGTGCCCCTCTATGACCGGAATCACGTCACTGGGGCCTCATCC
AGCAGCTCGTCCAGCACAAAGGCCACACTATATCCGCCGATCCTGAACCCACC
CCCGTCCCCGGCCACAGACCCCTCTCTCTACAACGTGGACGTGTTTTATTCTT
CAGGCATCCCGGCCACCGCTAGACCATACAGGCCCTACGTCATTCGAGGTATG
GCACCCCCAACAACACCGTGCAGCACAGATGTGTGTGACAGTGACTACAGCAT
CAGTCGCTGGAAGAGCAGCAAATACTACCTGGACTTGAATTCGGACTCAGACC
CCTACCCCCCCCCGCCCACCCCCCACAGCCAGTACCTATCTGCAGAGGACAGC
TGCCCACCCTCACCAGGCACTGAGAGGAGTTACTGCCACCTCTTCCCGCCCCC
ACCGTCCCCCTGCACGGACTCGTCCTGACCTCGGCCGTCCACCCGGCCCTGCT
GCCTCCCTGTAAATATTTTTAAATATGAACAAAGGAAAAATATATTTTATGAT
TTAAAAAATAAATATAATTGGGGTTTTTAACAAGTGAGAAATGTGAGCGGTGA
AGGGGTGGGCAGGGCTGGGAAACTTTTCTAG (SEQ ID NO:7)

FIG. 7C

METAPTRAPPPPPPPPLLLLVLYCSLVPAAASPLLLFANRRDVRLVDAGGVKLE
STIVASGLEDAAAVDFQFSKGAVYWTDVSEEAIKQTYLNQTGAAAQNIVISGL
VSPDGLACDWVGKKLYWTDSETNRIEVANLNGTSRKVLFWQDLDQPRAIALDP
AHGYMYWTDWGEAPRIERAGMDGSTRKIIVDSDIYWPNGLTIDLEEQKLYWAD
AKLSFIHRANLDGSFRQKVVEGSLTHPFALTLSGDTLYWTDWQTRSIHACNKW
TGEQRKEILSALYSPMDIQVLSQERQPPFHTPCEEDNGGCSHLCLLSPREPFY
SCACPTGVQLQDNGKTCKTGAEEVLLLARRTDLRRISLDTPDFTDIVLQVGDI
RHAIAIDYDPLEGYVYWTDDEVRAIRRAYLDGSGAQTLVNTEINDPDGIAVDW
VARNLYWTDTGTDRIEVTRLNGTSRKILVSEDLDEPRAIVLHPVMGLMYWTDW
GENPKIECANLDGRDRHVLVNTSLGWPNGLALDLQEGKLYWGDAKTDKIEVIN
IDGTKRKTLLEDKLPHIFGFTLLGDFIYWTDWQRRSIERVHKVKASRDVIIDQ
LPDLMGLKAVNVAKVVGTNPCADGNGGCSHLCFFTPRATKCGCPIGLELLSDM
KTCIIPEAFLVFTSRATIHRISLETNNNDVAIPLTGVKEASALDFDVSNNHIY
WTDVSLKTISRAFMNGSSVEHVIEFGLDYPEGMAVDWMGKNLYWADTGTNRIE
VARLDGQFRQVLVWRDLDNPRSLALDPTKGYIYWTEWGGKPRIVRAFMDGTNC
MTLVDKVGRANDLTIDYADQRLYWTDLDTNMIESSNMLGQERMVIADDLPYPF
GLTQYSDYIYWTDWNLHSIERADKTSGRNRTLIQGHLDFVMDILVFHSSRQDG
LNDCVHSNGQCGQLCLAIPGGHRCGCASHYTLDPSSRNCSPPSTFLLFSQKFA
ISRMIPDDQLSPDLVLPLHGLRNVKAINYDPLDKFIYWVDGRQNIKRAKDDGT
QPSMLTSPSQSLSPDRQPHDLSIDIYSRTLFWTCEATNTINVHRLDGDAMGVV
LRGDRDKPRAIAVNAERGYMYFTNMQDHAAKIERASLDGTEREVLFTTGLIRP
VALVVDNALGKLFWVDADLKRIESCDLSGANRLTLEDANIVQPVGLTVLGRHL
YWIDRQQQMIERVEKTTGDKRTRVQGRVTHLTGIHAVEEVSLEEFSAHPCARD
NGGCSHICIAKGDGTPRCSCPVHLVLLQNLLTCGEPPTCSPDQFACTTGEIDC
IPGAWRCDGFPECADQSDEEGCPVCSASQFPCARGQCVDLRLRCDGEADCQDR
SDEANCDAVCLPNQFRCTSGQCVLIKQQCDSFPDCADGSDELMCEINKPPSDD
IPAHSSAIGPVIGIILSLFVMGGVYFVCQRVMCQRYTGASGPFPHEYVGGAPH
VPLNFIAPGGSQHGPFPGIPCSKSVMSSMSLVGGRGSVPLYDRNHVTGASSSS
SSSTKATLYPPILNPPPSPATDPSLYNVDVFYSSGIPATARPYRPYVIRGMAP
PTTPCSTDVCDSDYSISRWKSSKYYLDLNSDSDPYPPPPTPHSQYLSAEDSCP
PSPGTERSYCHLFPPPPSPCTDSS (SEQ ID NO:8)

FIG. 8

Construct

Gene: 193   **GI Number(s):** 6678715

Gene Family: EGF domain protein

Gene Subfamily: Low-density lipoprotein receptor

Gene Sequence: full-length cDNA, Mouse

underlined = deleted in targeting construct

[ ] = sequence flanking Neo insert in targeting construct

```
GCCGCGGCGCCCGAGGCGGGAGCAAGAGGCGCCGGGAGCCGCGAGGATCCACCGCCGCCG
CGCGCGCCATGGAGCCCGAGTGAGCGCGCGGCGCTCCCGGCCGCCGGACGACATGGAAAC
GGCGCCGACCCGGGCCCCTCCGCCGCCGCCGCCGCCGCTGCTGCTGCTGGTGCTGTACTG
CAGCTTGGTCCCCGCCGCGGCCTCACCGCTCCTGTTGTTTGCCAACCGCCGGGATGTGCG
GCTAGTGGATGCCGGCGGAGTGAAGCTGGAGTCCACCATTGTGGCCAGTGGCCTGGAGGA
TGCAGCTGCTGTAGACTTCCAGTTCTCCAAGGGTGCTGTGTACTGGACAGATGTGAGCGA
GGAGGCCATCAAACAGACCTACCTGAACCAGAACTGGAGCTGCTGCACAGAACATTGTCAT
CTCGGGCCTCGTGTCACCTGATGGCCTGGCCTGTGACTGGGTTGGCAAGAAGCTGTACTG
GACGGACTCCGAGACCAACCGCATTGAGGTTGCCAACCTCAATGGGACGTCCCGTAAGGT
TCTCTTCTGGCAGGACCTGGACCAGCCAAGGGCCATTGCCCTGGATCCTGCACATGGGTA
CATGTACTGGACTGACTGGGGGGAAGCACCCCGGATCGAGCGGGCAGGGATGGATGGCAG
TACCCGGAAGATCATTGTAGACTCCGACATTTACTGGCCCAATGGGCTGACCATCGACCT
GGAGGAACAGAAGCTGTACTGGGCCGATGCCAAGCTCAGCTTCATCCACCGTGCCAACCT
GGACGGCTCCTTCCGGCAGAAGGTGGTGGAGGGCAGCCTCACTCACCCCTTTTGCCCTGAC
ACTCTCTGGGGACACACTCTACTGGACAGACTGGCAGACCCGCTCCATCCACGCCTGCAA
CAAGTGGACAGGGGAGCAGAGGAAGGAGATCCTTAGTGCTCTGTACTCACCCATGGACAT
CCAAGTGCTGAGCCAGGAGCGGCAGCCTCCCTTCCACACACCATGCGAGGAGGACAACGG
TGGCTGTTCCCACCTGTGCCTGCTGTCCCCGAGGGGAGCCTTTCTACTCCTGTGCCTGCCC
CACTGGTGTGCAGTTGCAGGACAATGGCAAGACGTGCAAGACAGGGGCTGAGGAAGTGCT
GCTGCTGGCTCGGAGGACAGACCTGAGGAGGATCTCTCTGGACACCCCTGACTTCACAGA
CATAGTGCTGCAGGTGGGCGACATCCGGCATGCCATTGCCATTGACTACGATCCCCTGGA
GGGCTACGTGTACTGGACCGATGATGAGGTGCGGGCTATCCGCAGGGCGTACCTAGATGG
CTCAGGTGCGCAGACACTTGTGAACACTGAGATCAATGACCCCGATGGCATTGCTGTGGA
CTGGGTCGCCCGGAACCTCTACTGGACAGATACAGGCACTGACAGAATTGAGGTGACTCG
CCTCAACGGCACCTCCCGAAAGATCCTGGTATCTGAGGACCTGGACGAACCGCGAGCCAT
TGTGTTGCACCCTGTGATGGGCCTCATGTACTGGACAGACTGGGGGGAGGAGAACCCCAAAAT
CGAATGCGCCAACCTAGATGGGGAGAGATCGGCATGTCCTGGTGAACACCTCCCTTGGGTG
GCCCAATGGACTGGCCCTGGACCTGCAGGAGGGCAAGCTGTACTGGGGGGGATGCCAAAAC
TGATAAAATCGAGGTGATCAACATAGACGGGACAAAGCGGAAGACCCTGCTTGAGGACAA
GCTCCCACACATTTTTTGGGTTCACACTGCTGGGGGACTTCATCTACTGGACCGACTGGCA
GAGACGCAGTATTGAAAAGGGTCCACAAGGTCAAGGCCAGCCGGGATGTCATCATTGATCA
ACTCCCCGACCTGATGGGACTCAAAGCCGTGAATGTGGCCAAGGTTGTCGGAACCAACCC
ATGTGCGGATGGAAATGGAGGGTGCAGCCATCTGTGCTTCTTCACCCCACGTGCCACCAA
GTGTGGCTGCCCCATTGGCCTGGAGCTGTTGAGTGACATGAAGACCTGCATAATCCCCGA
GGCCTTCCTGGTATTCACCAGCAGAGCCACCATCCACAGGATCTCCCTGGAGACTAACAA
CAACGATGTGGCTATCCCACTCACGGGTGTCAAAGAGGCCTCTGCACTGGACTTTGATGT
GTCCAACAATCACATCTACTGGACTGATGTTAGCCTCAAGACGATCAGCCGAGCCTTCAT
GAATGGGAGCTCAGTGGAGCACGTGATTGAGTTTGGCCTCGACTACCCTGAAGGAATGGC
TGTGGACTGGATGGGCAAGAACCTCTATTGGGCGGACACAGGGACCAACAGGATTGAGGT
GGCCCGGCTGGATGGGCAGTTCCGGCAGGTGCTTGTGTGGAGAGACCTTGACAACCCCAG
GTCTCTGGCTCTGGATCCTACTAAAGGCTACATCTACTGGACTGAGTGGGGTGGCAAGCC
AAGGATTGTGCGGGCCTTCATGGATGGGACCAATTGTATGACACTGGTAGACAAGGTGGG
CCGGGCCAACGACCTCACCATTGATTATGCCGACCAGCGACTGTACTGGACTGACCTGGA
CACCAACATGATTGAGTCTTCCAACATGCTGGGTCAGGAGCGCATGGTGATAGCTGACGA
```

```
TCTGCCCTACCCGTTTGGCCTGACTCAATATAGCGATTACATCTACTGGACTGACTGGAA
CCTGCATAGCATTGAACGGGCGGACAAGACCAGTGGGCGGAACCGCACCCTCATCCAGGG
TCACCTGGACTTCGTCATGGACATCCTGGTGTTCCACTCCTCCCGTCAGGATGGCCTCAA
CGACTGCGTGCACAGCAATGGCCAGTGTGGGCAGCTGTGCCTCGCCATCCCCGGAGGCCA
CCGCTGTGGCTGTGCTTCACACTACACGCTGGACCCCAGCAGCCGCAACTGCAGCCCGCC
CTCCACCTTCTTGCTGTTCAGCCAGAAATTTGCCATCAGCCGGATGATCCCCGATGACCA
GCTCAGCCCGGACCTTGTCCTACCCCTTCATGGGCTGAGGAACGTCAAAGCCATCAACTA
TGACCCGCTGGACAAGTTCATCTACTGGGTGGACGGGCGCCAGAACATCAAGAGGGCCAA
GGACGACGGTACCCAGCCCTCCATGCTGACCTCTCCCAGCCAAAGCCTGAGCCCAGACAG
ACAGCCACACGACCTCAGCATTGACATCTACAGCCGGACACTGTTCTGGACCTGTGAGGC
CACCAACACTATCAATGTCCACCGGCTGGATGGGGATGCCATGGGAGTGGGTGCTTCGAGG
GGACCGTGACAAGCCAAGGGCCATTGCTGTCAATGCTGAGCGAGGGTACATGTACTTTAC
CAACATGCAGGACCATGCTGCCAAGATCGAGCGAGCCTCCCTGGATGGCACAGAGCGGGA
GGTCCTCTTCACCACAGGCCTCATCCGTCCCGTGGCCCTTGTGGTGGACAATGCTCTGGG
CAAGCTCTTCTGGGTGGATGCCGACCTAAAGCGAATCGAAAGCTGTGACCTCTCTG[GGG
CCAACCGCCTGACCCTGGAAGATGCCAACATCGTACAGCCAGTAGGTCTGACAGTGCTGG
GCAGGCACCTCTACTGGATCGACCGCCAGCAGCAGATGATCGAGCGCGTGGAGAAGACC]
ACTGGGGACAAGCGGACTAGGGTTCAGGGCCGTGTCACCCACC[TGACAGGCATCCATGC
CGTGGAGGAAGTCAGCCTGGAGGAGTTCT]CAGCCCATCCTTGTGCCCGAGACAATGGCG
GCTGCTCCCACATCTGTATCGCCAAGGGTGATGGAACACCGCGCTGCTCGTGCCCTGTCC
ACCTGGTGCTCCTGCAGAACCTGCTGACTTGTGGTGAGCCTCCTACCTGCTCCCCTGATC
AGTTTGCATGTACCACTGGTGAGATCGACTGCATCCCCGGAGCCTGGCGCTGTGACGGCT
TCCCTGAGTGTGCTGACCAGAGTGATGAAGAAGGCTGCCCAGTGTGCTCCGCCTCTCAGT
TCCCCTGCGCTCGAGGCCAGTGTGTGGACCTGCGGTTACGCTGCGACGGTGAGGCCGACT
GCCAGGATCGCTCTGATGAAGCTAACTGCGATGCTGTCTGTCTGCCCAATCAGTTCCGGT
GCACCAGCGGCCAGTGTGTCCTCATCAAGCAACAGTGTGACTCCTTCCCCGACTGTGCTG
ATGGGTCTGATGAGCTCATGTGTGTGAAATCAACAAGCCACCCTCTGATGACATCCCAGCCC
ACAGCAGTGCCATTGGGCCCGTCATTGGTATCATCCTCTCCCTCTTCGTCATGGGCGGGG
TCTACTTTGTCTGCCAGCGTGTGATGTGCCAGCGCTACACAGGGGCCAGTGGGCCCTTTC
CCCACGAGTATGTTGGTGGAGCCCCTCATGTGCCTCTCAACTTCATAGCCCCAGGTGGCT
CACAGCACGGTCCCTTCCCAGGCATCCCGTGCAGCAAGTCCGTGATGAGCTCCATGAGCC
TGGTGGGGGGGCGCGGCAGCGTGCCCCTCTATGACCGGAATCACGTCACTGGGGCCTCAT
CCAGCAGCTCGTCCAGCACAAAGGCCACACTATATCCGCCGATCCTGAACCCACCCCCGT
CCCCGGCCACAGACCCCTCTCTCTACAACGTGGACGTGTTTTATTCTTCAGGCATCCCGG
CCACCGCTAGACCATACAGGCCCTACGTCATTCGAGGTATGGCACCCCCAACAACACCGT
GCAGCACAGATGTGTGTGACAGTGACTACAGCATCAGTCGCTGGAAGAGCAGCAAATACT
ACCTGGACTTGAATTCGGACTCAGACCCCTACCCCCCCCCGCCCACCCCCCACAGCCAGT
ACCTATCTGCAGAGGACAGCTGCCCACCCTCACCAGGCACTGAGAGGAGTTACTGCCACC
TCTTCCCGCCCCCACCGTCCCCCTGCACGGACTCGTCCTGACCTCGGCCGTCCACCCGGC
CCTGCTGCCTCCCTGTAAATATTTTTAAAATATGAACAAAGGAAAAATATATTTTATGATT
TAAAAAATAAATATAATTGGGGTTTTTAACAAGTGAGAAATGTGAGCGGTGAAGGGGTGG
GCAGGGCTGGGAAACTTTTCTAG
```

Gene Sequence Structure *	3659 bp	Sequence Deleted	3701 bp

Size of full-length
cDNA: 5119 bp

FIG. 9B

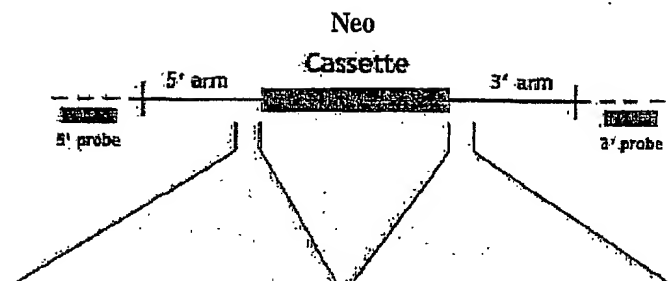Targeting Vector* (genomic sequence)

Construct Number: 992

Arm Length:
5': 1.5 kb
3': 2.9 kb

———— Targeting Vector
— — — — Endogenous Locus

* Not drawn to scale


```
5'>AAATATGCATTATCCTGAGCA      5'>TCACTGGCATCCATGCAGTG:
CAGTGGGTCTGGCCCTTCACTTGG      AGGAAGTCAGCCTGGAGGAGTTCT
CTGCCACTCATGGAGCCTTTATGC      GTACGTGAGAGGGGACAGTGTTTG
TAACCACAGGGGCCAACCGCCTGA      TGGTGGGGTCTCCTGGGGGAAGGT
CCCTGGAAGATGCCAACATCGTAC      GAATCAGCCCTACTGGCATCAGAT
AGCCAGTAGGTCTGACAGTGCTGG      GGGCTGCTGGTGCAAGAGCAGTGT
GCAGGCACCTCTACTGGATCGACC      GCCTGAGGAGCTCATGGGCTCAGC
GCCAGCAGCAGATGATCGAGCGTG      ACCGAAGGCCAGTGCATGTCCAGA
TGGAGAAGACC<3'               TGTCTGCCTCT<3'
(SEQ ID NO:9)                (SEQ ID NO:10)
```

FIG. 9C

Phenotypic Data Summary - Open Field

Legend:
- 2,1,-/-,Male
- 2,1,+/+,Male

Y-axis: Average session time in central (%)
X-axis: Gene
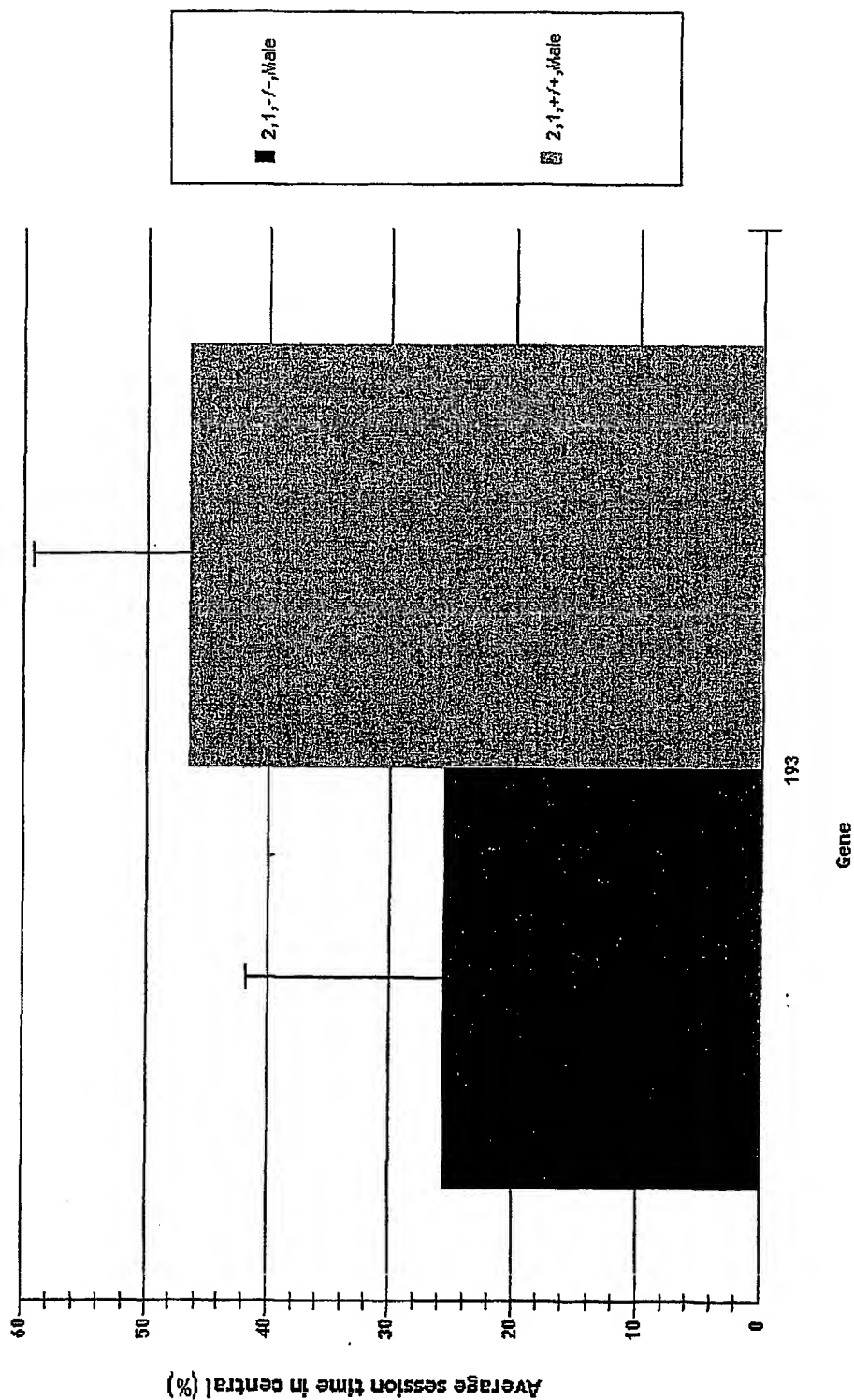
FIG. 10

Phenotypic Data Summary - Open Field

FIG. 11

Legend:
- 2,1,-/-,Male
- 2,1,+/+,Male

```
GGGGGGGGGGGGGGGTCAGAGGGAGCTTTCTTTTAGGCCCGTCCATCTGTGAAT
CTAACCTCAGTTTCTGGGAATCAGGAAGCATGCATCTCCTCTTAGTTCAGCTG
CTTGTTCTCTTGCCTCTGGGGAAGGCAGACCTATGTGTGGATGGCTGCCAGAG
TCAGGGCTCTTTATCCTTTCCTCTCCTAGAAAGGGGTCGCAGAGATCTCCACG
TGGCCAACCACGAGGAGGCAGAAGACAAGCCGGATCTGTTTGTGGCCGTGCCA
CACCTCATGGGCACCAGCCTGGCTGGGGAAGGCCAGAGGCAGAGAGGGAAGAT
GCTGTCCAGGCTTGGAAGATTCTGGAAGAAACCTGAGACCGAATTTTACCCCC
CAAGGGATGTGGAAAGCGATCATGTCTCATCGGGGATGCAGGCCGTGACTCAG
CCAGCAGATGGGAGGAAAGTGGAGAGATCACCTCTACAGGAGGAAGCCAAGAG
GTTCTGGCATCGGTTCATGTTCAGAAAGGGCCCGGCGTTCCAGGGAGTCATCC
TGCCCATCAAAAGCCACGAAGTACACTGGGAGACCTGCAGGACTGTGCCCTTC
AACCAGACCATTGCCCATGAAGACTGTCAAAAAGTCGTTGTCCAGAACAACCT
TTGCTTTGGCAAATGCAGTTCCATTCGTTTTCCCGGAGAAGGGGCAGATGCCC
ACAGCTTCTGCTCCCACTGCTCGCCCACCAAATTCACCACCGTGCACTTGATG
CTGAACTGCACCAGCCCAACCCCCGTGGTCAAGATGGTGATGCAAGTAGAAGA
GTGTCAGTGCATGGTGAAGACGGAACGTGGAGAGGAGCGCCTCCTACTGGCTG
GTTCCAGGGTTCCTTCATCCCTGGACTTCCAGCTTCAAAAACAAACCCATGA
ATTACCTCAACAGAAAGCAAAACCTCAACAGAATAAGTGAGGGTTATTCAATC
TGGAAATGTTATGTGAGTTATATAAAGATCAGTGGAAAATATCTTTCTCTCTC
CCTCTCTCCCCCTCTCTCTTCTCTCTATTTTCTCTCTCTCTCTCTCTCTCT
CTCTCTCTCTCTCTCTCTCACACACACACACACACACACACACACACACACAT
GTTTGTGTTTAGACAGGGTCTTATGTATTCTCAGCTGGCCTCAAACTCACAAT
GTGGCTGGGGATGATTTTAAACTCCTGATCCAATTCCTGAGTGCTGGGATTAC
AGACATGCTCCATAANACATAGCTCCCAGAAGGATTTTTAAAAGAGATTTTGC
ATGTTTCAAAGTTGCCTTTGAGACTCAGAAATATTTTGATNTATTGAATGGCC
TTGCCACAGATGTGGGAGGCAGCTTGCTTGGTGGCCCAAGTATTTTTTTTTTG
TTCGTTCAGAATTCTCCACATGAAGTTTTTACTGTTGGTTATCTGGCGTTGAA
GAAGGAATAGTGAAGGTACTTTTAACAGTTTACACGTGGAAGGGGCTCAGGCA
CTAGGAACCAACCTTTTCCCGGAATATGAGGAAAATACATGAACAGTATTAGA
GTCACTTGAGGAAGTTACTAGGAAACGCCATAAGTCTCCAAGTACATTGTGAG
TCATTTTGAAGGACAATCGTGTATATAGACGAAATCTTCTACTCGTATGCTTT
TGAATCTTCTAGCAAGTTAGGTTTCTATGTTTGGGCTTCTTCCTATTGTCTAA
GAGTATGTGTGACAAATTCAACCTGACAAATACCTCAATGGCAAATTCTGACC
CTG (SEQ ID NO:11)
```
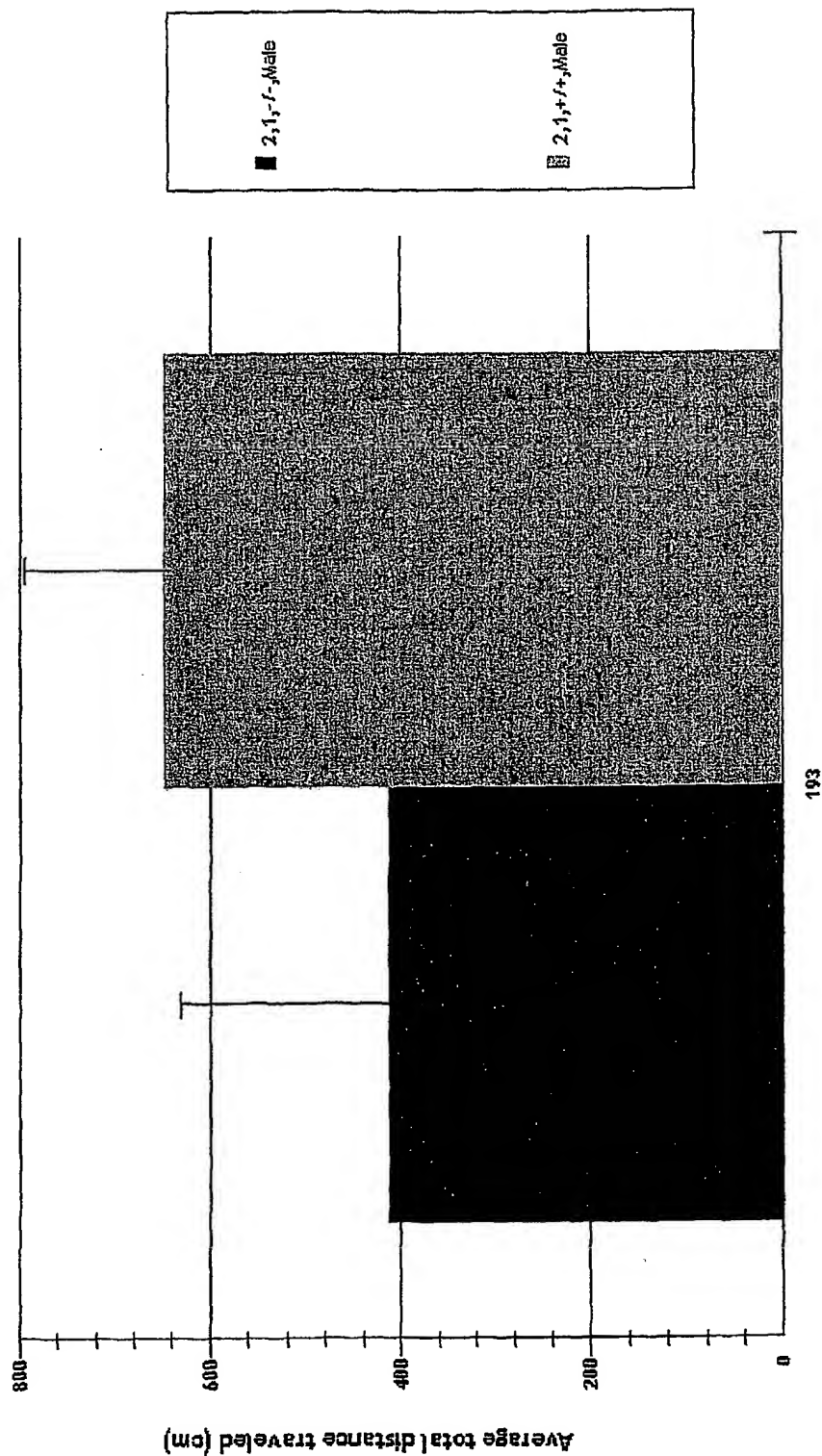
```
MHLLLVQLLVLLPLGKADLCVDGCQSQGSLSFPLLERGRRDLHVANHEEAEDK
PDLFVAVPHLMGTSLAGEGQRQRGKMLSRLGRFWKKPETEFYPPRDVESDHVS
SGMQAVTQPADGRKVERSPLQEEAKRFWHRFMFRKGPAFQGVILPIKSHEVHW
ETCRTVPFNQTIAHEDCQKVVVQNNLCFGKCSSIRFPGEGADAHSFCSHCSPT
KFTTVHLMLNCTSPTPVVKMVMQVEECQCMVKTERGEERLLLAGSQGSFIPGL
PASKTNP (SEQ ID NO:12)
```

FIG. 12

Construct

Gene: 67   **GI Number(s):** 6753409

Gene Family: Growth Factor Inhibitor

Gene Subfamily: Cerberus

Gene Sequence: full-length cDNA, Mouse

underlined = deleted in targeting construct

[ ] = sequence flanking Neo insert in targeting construct

```
[GGGGGGGGGGGGGGGTCAGAGGGAGCTTTCTTTTAGGCCCGTCCATCTGTGAATCTAACC
TCAGTTTCTGGGAATCAGGAAGCATGCATCTCCTCTTAGTTCAGCTGCTTGTTCTCTTGC
CTCTGGGGAAGGCAGACCTATGTGTGGATGGCTGCCAGAGTCAGGGCTCTTTATCCTTTC
CCTCCTAGAAAGGGGTCGCAGAGATCTCCACGTGGCCAACCACGAGGAGGCAGAAGACAA
]GCCGGATCTGTTTGTGGCCGTGCCACACCTCATGGGCACCAGCCTGGCTGGGGAAGGCC
AGAGGCAGAGAGGGAAGATGCTGTCCAGGCTTGGAAGATTCTGGAAGAAACCTGAGACCG
AATTTTACCCCCCAAGGGATGTGGAAAGCGATCATGTCTCATCGGGGATGCAGGCCGTGA
CTCAGCCAGCAGATGGGAGGAAAGTGGAGAGATCACCTCTACAGGAGGAAGCCAAGAGGT
TCTGGCATCGGTTCATGTTCAGAAAGGGCCCGGCGTTCCAGGGAGTCAT[CCTGCCCATC
AAAAGCCACGAAGTACACTGGGAGACCTGCAGGACTGTGCCCTTCAACCAG]ACCATTGC
CCATGAAGACTGTCAAAAAGTCGTTGTCCAGAACAACCTTTGCTTTGGCAAATGCAGTTC
CATTCGTTTTCCCGGAGAAGGGGCAGATGCCCACAGCTTTCTGCTCCCACTGCTCGCCCAC
CAAATTCACCACCGTGCACTTGATGCTGAACTGCACCAGCCCAACCCCCGTGGTCAAGAT
GGTGATGCAAGTAGAAGAGTGTCAGTGCATGGTGAAGACGGAACGTGGAGAGGAGCGCCT
CCTACTGGCTGGTTCCCAGGGTTCCTTCATCCCTGGACTTCCAGCTTCAAAAACAAACCC
ATGAATTACCTCAACAGAAAGCAAAACCTCAACAGAATAAGTGAGGGTTATTCAATCTGG
AAATGTTATGTGAGTTATATAAAGATCAGTGGAAAATATCTTTCTCTCTCCCTCTCTCCC
CCTCTCTCTTCTCTCTATTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC
TCACACACACACACACACACACACACACATGTTTGTGTTTAGACAGGGTCTTAT
GTATTCTCAGCTGGCCTCAAACTCACAATGTGGCTGGGGATGATTTTAAACTCCTGATCC
AATTCCTGAGTGCTGGGATTACAGACATGCTCCATAANACATAGCTCCCAGAAGGATTTT
TAAAAGAGATTTTGCATGTTTCAAAGTTGCCTTTGAGACTCAGAAATATTTTGATNTATT
GAATGGCCTTGCCACAGATGTGGGAGGCAGCTTGCTTGGTGGCCCAAGTATTTTTTTTTT
GTTCGTTCAGAATTCTCCACATGAAGTTTTTACTGTTGGTTATCTGGCGTTGAAGAAGGA
ATAGTGAAGGTACTTTTAACAGTTTACACGTGGAAGGGGCTCAGGCACTAGGAACCAACC
TTTTCCCGGAATATGAGGAAAATACATGAACAGTATTAGAGTCACTTGAGGAAGTTACTA
GGAAACGCCATAAGTCTCCAAGTACATTGTGAGTCATTTTGAAGGACAATCGTGTATATA
GACGAAATCTTCTACTCGTATGCTTTTGAATCTTCTAGCAAGTTAGGTTTCTATGTTTGG
GCTTCTTCCTATTGTCTAAGAGTATGTGTGACAAATTCAACCTGACAAATACCTCAATGG
CAAATTCTGACCCTG
```

Gene Sequence Structure *	241 bp	Sequence Deleted	528 bp

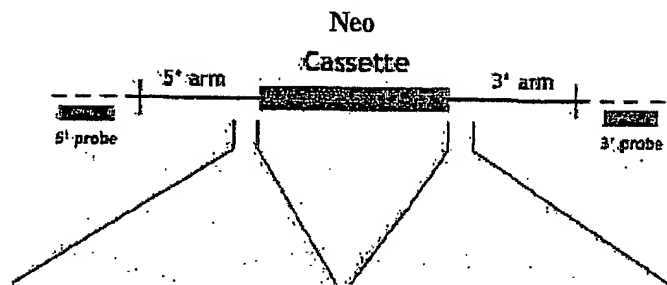**Size of full-length cDNA:** 1752 bp

FIG. 13A

Targeting Vector* (genomic sequence)

Construct Number: 269

Arm Length:
5': 2 kb
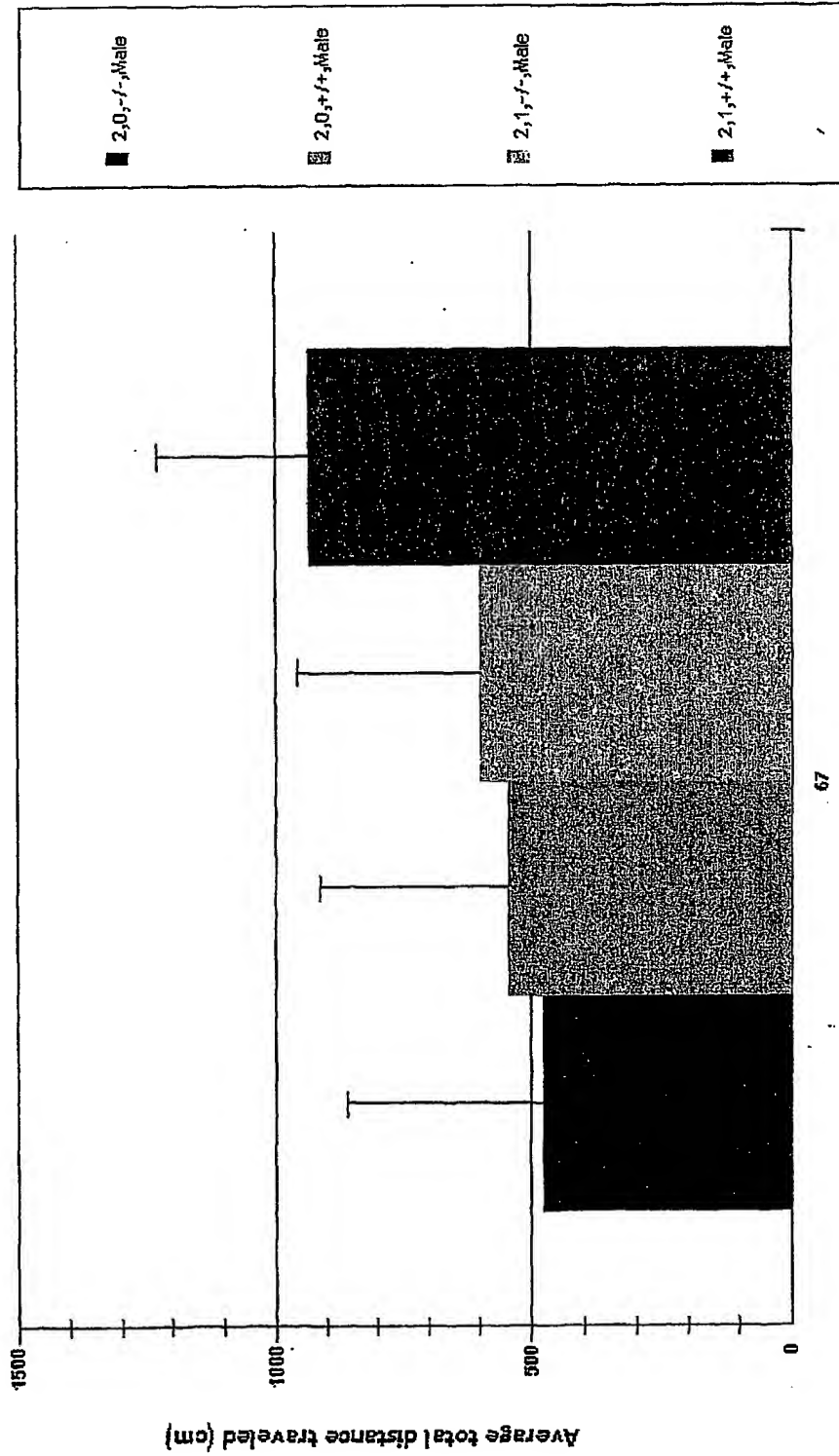3': 1.7 kb

———————— Targeting Vector
– – – – – Endogenous Locus

* Not drawn to scale

5'>TCCATCTGTGAATCTAACCTC
AGTCTCTGGGAATCAGGAAGCATG
CATCTCCTCTTAGTTCAGCTGCTT
GTTCTCTTGCCTCTGGGGAAGGCA
GACCTATGTGTGGATGGCTGCCAG
AGTCAGGGCTCTTTATCCTTTCCT
CTCCCAGAAAGGGGTCGCAGAGAT
CTCCACGTGGCCAACCACGAGGAG
GCAGAAGACAA<3'
(SEQ ID NO:13)

5'>CCTGCCCATCAAAAGCCACGA
AGTACACTGGGAGACCTGCAGGAC
TGTGCCCTTCAACCAGGTATGCAT
TCTAGAGGGTAAACCACCAGTTTG
CCAGACAGGGAGGACAGCTGGACA
GCTAGGACAAACGGCAAAATAGAA
AGAGTCTGGCGAGAGCTCGGGCCT
TGTCTAGTTCCAGATTCAGTCCTT
TGGGTATTCAT<3'
(SEQ ID NO:14)

FIG. 13B

Phenotypic Data Summary - Open Field

FIG. 14

Legend:
- 2,0,-/-,Male
- 2,0,+/+,Male
- 2,1,-/-,Male
- 2,1,+/+,Male

Y-axis: Average total distance traveled (cm) — 0, 500, 1000, 1500

X-axis: Gene — 67

FIG. 15

Phenotypic Data Summary - Tail Suspension

Legend:
- 2,0,-/-,Male
- 2,0,+/+,Male
- 2,1,-/-,Male
- 2,1,+/+,Male

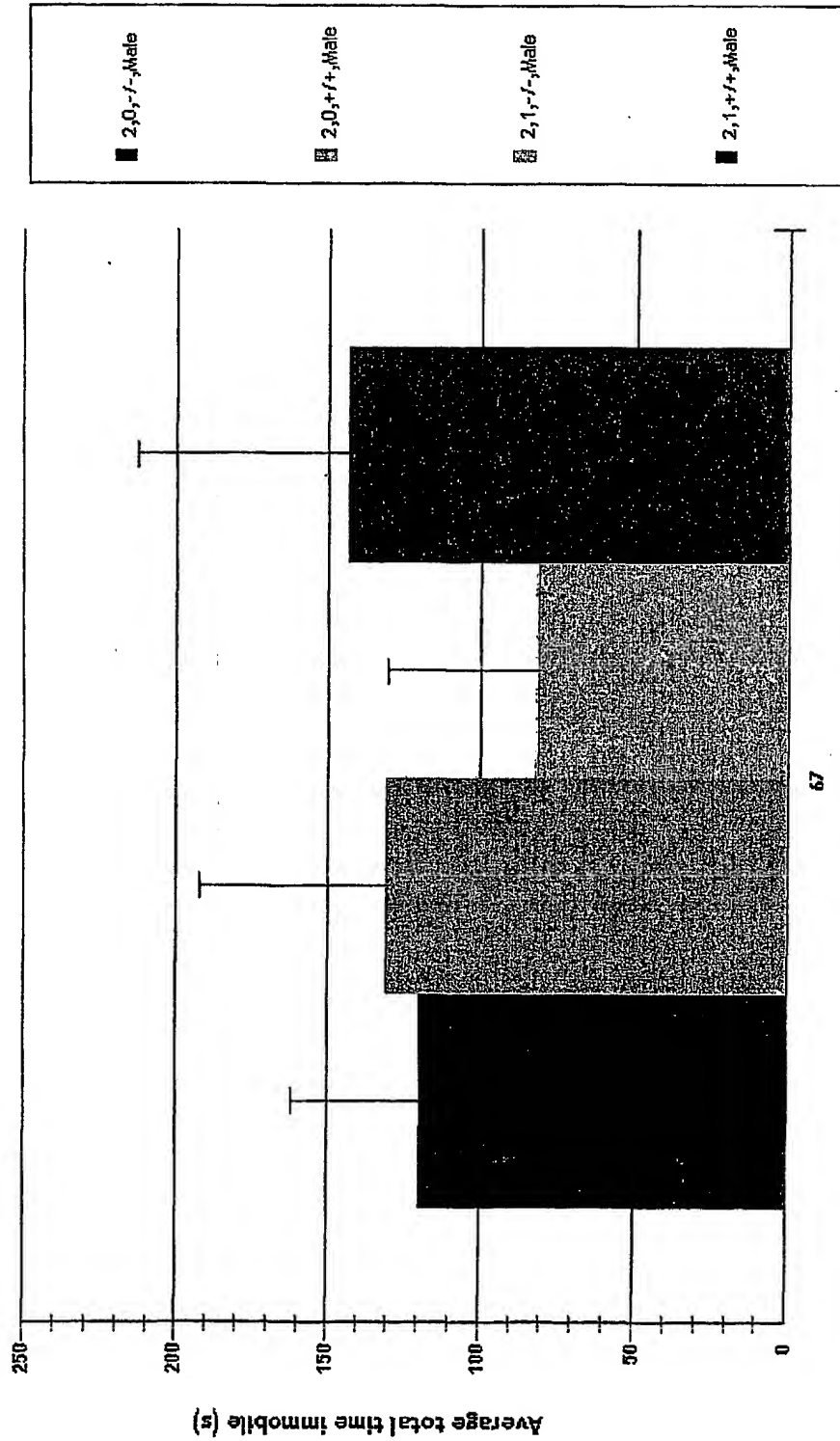Y-axis: Average total time immobile (s)
X-axis: Gene (67)

```
TCGGGGCTAGGGCTTGTGGGTAGCT (SEQ ID NO:15)
ACAATTCTGAAAACTGGTCCCTGGT (SEQ ID NO:16)
```

```
GTTCCTGGATGGCTGGGACCCANCCAGGACCTCCTGGTGGGAGATGTCATCCG
CCAAGCCCTGCCCAACAGGAACACTAAGGTCAAATACTTCATCCCACCCTCAA
TGTACAGGGCCACCCACTACCCACCCTATGCTGGTGGGGGAGGATATGTCATG
TCCAGAGCCACAGTGCGGCGCCTCCAGGCTATCATGGAAGATGCTGAACTCTT
CCCCATTGATGATGTCTTTGTGGGTATGTGCCTGAGGAGGCTGGGGCTGAGCC
TATTGCACCATGCTGGCTTCAAGACATTTGGAATCCGGCGGCCCCTGGACCCC
TTAGACCCCTGCCTGTATAGGGGGCTCCTGCTGGTTCACCGCCTCAGCCCCCT
CGAGATGTGGACCATGTGGGCACTGGTGACAGATGAGGGGCTCAAGTGTGCAG
CTGGCCCCATACCCCAGCGCTGAAGGGTGGGTTGGGCAACAGCTGAGAGTGGA
CTCAGTGTTGATTCTCTATCGTGATGCGAAATTGATGCCTGCTGCTCTACAGA
AAATGCCAACTGGGTTTTAACTCCTCTCACCNGTTAGNTCGATTANAACACTG
CAACCCAAAAAAAAAAAAAAAANCGGGGGGTTTTTTTGTCTTTTTTTTTTTTTT
TAGAGNGCGCAACGT (SEQ ID NO:17)
```

Figure 16